SARCASM DETECTION AND SURVEYING USER AFFECTATION

S. Maheswari*[1] and Dr. R. Anbuselvi[2]

[1]Research Scholar, Research and Development Department, Bharathiar University, Coimbatore.

[2]Assistant Professor, Department of Computer Science, Bishop Heber College, Trichy-17.

***Corresponding Author**
S. Maheswari
Research Scholar, Research
and Development
Department, Bharathiar
University, Coimbatore.

ABSTRACT

Sentiment Analysis is a technique to identify people's opinion, attitude, sentiment, and emotion towards any specific target such as individuals, events, topics, product, organizations, services etc. Sarcasm is a special kind of sentiment that comprise of words which mean the opposite of what you really want to say. Sarcasm is a sort of sentiment where public expresses their negative emotions using positive word within the text. It is very hard for humans to acknowledge. In this way we show the interest in sarcasm detection of social media text, particularly in tweets. In this paper we study new method pattern based approach for sarcasm detection, and also used behavioral modeling approach for effective sarcasm detection by analyzing the content of tweets however by conjoint exploiting the activity traits of users derived from their past activities. By using the various classifiers such as Random Forest, Support Vector Machine (SVM), k Nearest Neighbors (k-NN) and Maximum Entropy, we check the accuracy and performance.

KEYWORDS: Sarcasm, Patter Based Approach, User affection modeling, Sentiment analysis, SVM, KNN, Twitter.

1. INTRODUCTION

Sentiment analysis is the field of study that analyses people's sentiments, attitudes, and emotions from text. It is one of the most active research areas widely studied in data mining, Web mining, and text mining. Data mining refers to extracting knowledge from large

amounts of data.[11] One of the sub domain of data mining is Web Mining which extracts knowledge from the WWW.[11,12]

The web mining is divided in to three domains[11,12] which are as follows:

- Web Usage Mining.[12]
- Web Content Mining.[12]
- Web Structure Mining.[12]

Here for Sentiment Analysis the data of interest is only the text data, so Text mining is done on the content of the web. There are many challenges in Sentiment Analysis and one of them is sarcasm detection. Sentiment analysis can be easily misled by the presence of words that have a strong polarity but are used sarcastically, which means that the opposite polarity was intended.

Social net-working websites have become a popular platform for users to express their feelings and opinions on various topics, such as events, or products. Social media channels have become a popular platform to discuss ideas and to interact with people worldwide area. Twitter is also important social media network for people to express their feelings, opinions, and thoughts. Users post more than 340 million tweets and 1.6 billion search queries every day.[4,3]

Twitter is a social media platform where users post their views of everyday life. Many organizations and companies have been interested in these data for the purpose of studying the opinion of people regards the political events, popular products or Movies. When a particular product is launched, people start tweeting, writing reviews, posting comments, etc. on social media such as twitter. People turn to social media network to read the comments, and reviews from other users about a product before they decide whether to purchase or not. If the user review is good for the particular products then the users are buy the product otherwise not. Organizations are also depends on these sites to know the response of users for their products and use the user feedback to improve their products.[2] Sentiment analysis is the opinion of the user for the particular things. Sentiment analysis is the extraction of feeling from any communication (verbal/non verbal).Two ways to express sentiment analysis.

1) Explicit sentiments: Direct expression of the opinion about the subject shows the presence of explicit sentiment.

2) Implicit sentiments: Whenever any sentence implies an opinion then such sentence shows the Presence of implicit sentiment (Indirect expression).

Sentiment analysis and opinion mining depends on emotional words in a text to check its polarity (i.e., whether it deals positively or negatively with its theme).[5] Sarcasm is a type of sentiment where people express their negative feelings using positive word in the text.

2. LITERATURE REVIEW

In,[1] authors used the machine learning approach to sarcasm detection on Twitter in two languages English and Czech. First work is sarcasm detection on Czech language. They used the two classifier Maximum Entropy (MaxEnt) and Support Vector Machine (SVM) with different combinations of features on both the Czech and English datasets. Also use the different preprocessing technique such as Tokenizing, POS-tagging, No stemming and Removing stop words, its use for finding the issue of Czech language.

In,[9] authors have investigated characteristics of sarcasm on Twitter. They are concerned not just with identifying whether tweets are sarcastic or not, but also consider the polarity of the tweets. They also have compiled a number of rules which improve the accuracy of sentiment analysis when sarcasm is known to be present. Researcher have developed a hash tag tokenizes for GATE method so that sentiment and sarcasm found within hash tag can be detected more easily. Hash tag tokenization method is very useful for detection of sarcasm and checks the polarity of the tweet i.e. positive or negative.

In,[8] authors are used two methods such as lexical and pragmatic factors that are use for differentiate between sarcasm from positive and negative sentiments expressed in Twitter messages. They also created corpus of sarcastic Twitter messages in which determination of the sarcasm of each message has been made by its author. Corpus is used to compare sarcastic utterances in Twitter to utterances that show positive or negative attitudes without sarcasm.

In,[5] authors use the computational system it is use for harnesses context incongruity as a basis for sarcasm detection. Sarcasm classifier uses four types of features: lexical, pragmatic, explicit incongruity, and implicit incongruity features. They evaluate system on two text forms: tweets and discussion forum posts. For improvement of performance of tweet uses the rule base algorithm, and to improve the performance for discussion forum posts, uses the

novel approach to use elicitor posts for sarcasm detection. This system also introduces error analysis, the system future work (a) role of numbers for sarcasm, and (b) situations with subjective sentiment.

Rule-based approaches attempt to identify sarcasm through specific evidences. These evidences are captured in terms of rules that rely on indicators of sarcasm. Focus on identifying whether a given simile (of the form '* as a *') is intended to be sarcastic. They use Google search in order to determine how likely a simile is. They present a 9-step approach where at each step rule; a simile is validated using the number of search results. Strength of this approach is that they present an error analysis corresponding to multiple rules.[10] The hash tag sentiment is a key indicator of sarcasm. Hash tags are often used by tweet authors to highlight sarcasm, and hence, if the sentiment expressed by a hash tag does not agree with rest of the tweet, the tweet is predicted as sarcastic. They use a hash tag tokenizer to split hashtags made of concatenated words.[9]

Following are the method for sarcasm detection on twitter.

1) Feature extraction

This method are used for annotating the data, it contain three categories.

a) Sarcasm as wit: when used as a wit, sarcasm is used with the purpose of being funny.

b) Sarcasm as whimper: when used as whimper, sarcasm is employed to show how annoyed or angry the person is.

c) Sarcasm as evasion: it refers to the situation when the person wants to avoid giving a clear answer, thus, makes use of sarcasm.

2) Sentiment-related Features

It extracts sentimental components of the tweet and counts them. Positive emotional content (e.g. love, happy, etc.) and negative emotional content (e.g. hate, sad, etc.).Calculate the ratio of emotional words.

$$p\,(t) = (\& \cdot PW + pw) - (\& \cdot NW + nw)/\,(\& \cdot PW + pw) + (\& \cdot NW + nw)\ 1$$

t=tweet, pw=positive words, nw =negative words, PW=highly emotional positive words, NW= highly emotional negative words, & =weight bigger than 1.

3) Punctuation-Related Features

It displays behavioral aspects such as low tones, Facial gestures or exaggeration. These aspects are translated into a certain use of punctuation or repetition of vowels when the message is written.

- Number of exclamation marks
- Number of question marks
- Number of dots
- Number of all-capital words
- Number of quotes.

4) Syntactic and Semantic Features

It refers to the situation when the person wants to avoid giving a clear answer, thus, makes use of sarcasm.

- Use of uncommon words
- Number of uncommon words
- Existence of common sarcastic expressions
- Number of interjections
- Number of laughing expressions.

5) Pattern-Related Features

Pattern is defined as an order sequence of words. Divide words into two classes: a first one referred to as CI containing words of which the content is important and a second one referred to as GFI containing the words of which the grammatical function is more important.

6) A behavioural modelling approach

In this method content to study how to develop a systematic approach for effective sarcasm detection by not only analyzing the content of the tweets but by also exploiting the behavioral traits of users derived from their past activities.[6]

3. PATTERN BASED APPROACH

Sarcasm is a type of sentiment where public expresses their negative emotions using positive word in the text.[5] It is usually used to convey implicit information within the message a person transmits. Sarcasm may be used for different purposes such as criticism. It is very difficult for humans to understand. Recognizing sarcastic statements can be very useful to improve automatic sentiment analysis of data collected from different websites or social

networks. Sarcasm is when a person says something different from what he means. Pattern based approach is used for detecting sarcasm on twitter.
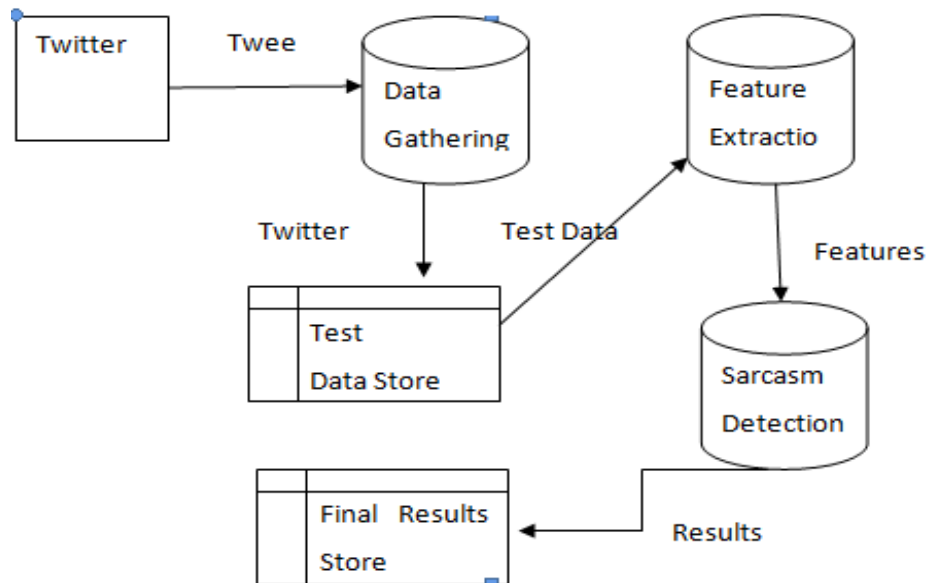

Fig 1: block diagram of sarcasm detection on twitter.

4. CONCLUSION

Sarcasm detection research has grown drastically in the past few years, demanding a look-back at the overall picture that these individual works have led to. This paper surveys approaches for automatic sarcasm detection. We have studied the different method for sarcasm detection; we also studied the pattern based approach for sarcasm detection. In this paper, the methods are used to detect sarcasm or as well as check the behavioral approach of the user, the method make used different component of the tweet, and also by using of Part-of-Speech tags to take out patterns characterizing the level of sarcasm of tweets. By using #sarcasm collect all the sarcastic tweets. In this way we discuss the different method such as Feature extraction, Sentiment-related Features, Punctuation-Related Features, Syntactic and Semantic Features, Pattern-Related Features, behavioral modeling approach for detection of sarcasm in the tweet. By using different algorithm or classifier such as Random Forest, Support Vector Machine (SVM), k Nearest Neighbors (k-NN) and Maximum Entropy, check the accuracy and performance. In future scope these approaches will show good results.

REFERENCES

1. Toma Ptacek Ivan Habernal and Jun Hong "Sarcasm Detection on Czech and English Twitter", Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, pages 213–223, Dublin, Ireland, August 23-29 2014.

2. S.K. Bharti B. Vachha, R.K. Pradhan, K.S. Babu, S.K. Jena "Sarcastic sentiment detection in tweets Streamed in real time: a big data approach", Elsevier 12 July 2016.

3. W.Tan, M.B.Blake, I.saleh, S.Dustdar, Social-network-sourced big data analytics, Internet Comput, 2013; 17(5): 62–69.

4. D. Chaffey, Global Social Media Research Summary 2016. URL ⟨http://www.smartinsights.com/Social-media-marketing/social-media-strategy/new-global-social-media-research/⟩.

5. Aditya Joshi, Vinita Sharma, Pushpak Bhattacharyya "Harnessing Context Incongruity for Sarcasm Detection" Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Short Papers), pages 757–762, Beijing, China, July 26-31, 2015. C 2015 Association for Computational Linguistic.

6. A. Rajadesingan, R. Zafarani, and H. Liu, ``Sarcasm detection on Twitter A behavioral modeling Approach", in Proc. 18th ACM Int. Conf. Web Search Data Mining, Feb. 2015; 79-106.

7. E. Riloff, A. Qadir, P. Surve, L. De Silva, N. Gilbert, and R. Huang, "Sarcasm as contrast between a positive Sentiment and negative situation", in Proc. Con Empirical Methods Natural Lang. Process, Oct. 2013; 704_714.

8. R. Gonzalez-Ibanez, S. Muresan, and N. Wacholder. 2011. "Identifying Sarcasm in Twitter: A Closer Look". In Proceedings of the 49th Annual Meeting of Association for Computational Linguistics.

9. D. Maynard, M. A. Greenwood. 2014. "Who cares about sarcastic tweets? Investigating the Impact of sarcasm on sentiment analysis", In Proceedings of the LREC, 2014 May 26-31.

10. Tony Veale and Yanfen Hao. "Detecting Ironic Intent in Creative Comparisons", In ECAI, 2010; 215: 765–770.

11. Data Mining Concepts and Techniques, J.Han M. Kamber.

12. G. Upadhyay K. Dhingra, Web Content Mining: Its Techniques and Uses IJARCSSE, 2013; 3(11).