Journal of Science and Technology
ISSN: 2456-5660 Volume 9, Issue 01 (Jan -2024)
www.jst.org.in *DOI:https://doi.org/10.46243/jst.2024.v9.i01.pp116-124*

SARCAMNET: EXTENSION OF LEXICON ALGORITHM FOR EMOJI-BASED SARCASM DETECTION FROM TWITTER DATA

Dr.SUBBA REDDY BORRA[1], T. KRUTHIKA[2], T.SAI POOJITHA[2], P. DEEPIKA[2]

[1]Professor and Head, Department of Information Technology, Mallareddy Engineering College for Women,(UGC-Autonomous), Hyderabad, India, bvsr79@gmail.com.

[2]Student, Department of Information Technology, Mallareddy Engineering College for Women, (UGC-Autonomous), Hyderabad, India.

To Cite this Article
Dr.SUBBA REDDY BORRA, T. KRUTHIKA, T.SAI POOJITHA, P. DEEPIKA ,**"SARCAMNET: EXTENSION OF LEXICON ALGORITHM FOR EMOJI-BASED SARCASM FROM DATA"** *Journal of Science and Technology, Vol. 09, Issue 01,-Jan 2024, pp116-124*

Article Info
Received: 29-12-2023 Revised: 09-01-2024 Accepted: 19-01-2024 Published: 29-01-2024

ABSTRACT :-

Lexicon algorithm is used to determine the sentiment expressed by a textual content. This sentiment might be negative, neutral, or positive. It is possible to be sarcastic using only positive or neutral sentiment textual contents. Hence, lexicon algorithm can be useful but insufficient for sarcasm detection. It is necessary to extend the lexicon algorithm to come up with systems that would be proven efficient for sarcasm detection on neutral and positive sentiment textual contents. In this paper, two sarcasm analysis systems both obtained from the extension of the lexicon algorithm have been proposed for that sake. The first system consists of the combination of a lexicon algorithm and a pure sarcasm analysis algorithm. The second system consists of the combination of a lexicon algorithm and a sentiment prediction algorithm. Finally, naive bayes are used to predict sarcasm detection using pretrained features.

Keyterms: Lexicon, sentiment, textual contents,extension, combination.

Journal of Science and Technology
ISSN: 2456-5660 Volume 9, Issue 01 (Jan -2024)
www.jst.org.in *DOI:https://doi.org/10.46243/jst.2024.v9.i01.pp116-124*

1. INTRODUCTION

Communication is the process of exchanging information. As time goes by, many ways and platforms of communication are being developed. Since the industrial revolution, the original way of communicating; face-to-face communication has been used as a model to develop the various ways of communicating known to date. Transposing the principles and codes of natural face-to-face communication to today's online communication is a major challenge for developers. Sarcasm is the communication practice that consists of meaning the opposite of what is said to mock or insult someone [1]. Sarcasm makes use of positive lingual contents to convey a negative message. Different types of approaches have been developed to implement sarcasm detection on online communication platforms. However, the levels of efficiency of these approaches have been the principal worries of developers. In this paper, propositions are made on how the lexicon algorithm can be extended to come out with systems that would be proven more efficient for sarcasm detection on textual contents. The magnitude of data generated through social media today is colossal. They are good for data analysis since they are very personal [1]. For years, companies have been analyzing this type of data to leverage their position in the market of their choice [2]. This field is called sentiment analysis [3]. On the other hand, sarcasm is defined as a positive utterance or sentence with underlying negative intention [4]. It is regarded as one of the most challenging issues in the Natural Language Processing (NLP) field [5]. Spotting and handling them correctly is crucial in an automated NLP system, mainly since sarcasm can flip the polarity of a sentence [5], [6]. Traditional studies as from Davidov et al. [7] and Riloff et al. [4] used rule-based techniques to tackle sarcasm detection. However, more recent studies [8], [9] have shifted towards deep learning to automatically detect the discriminatory features. In this work, the features extracted by a deep learning architecture is combined with the ones that are manually created through specific contextual understanding and processes. Tweets are used as the main source of input. Unlike in writing, different tones and gestures can be utilized to portray sarcasm in the real world [7]. As a countermeasure for this short-coming, writers of tweets tend to leave contextual clues for sarcasm in creative ways such as hashtags and hyperboles [4], [7]. This kind of clue is what this work is

trying to find and exploit. Several NLP studies have tried to come up with automatic detection models for sarcasm. Features are either discovered through deep learning or manual handcrafting (feature engineering) methods [10], never both. There is too much reliance on deep learning architecture for some researchers [8], [9], and vice-versa for manual handcrafting. This leaves some room for experiments. The method also significantly improves the F1-measure from the existing study using the same dataset. This work also demonstrates the generality of a deep learning architecture. For future work, a few datasets will be used to further generalize the comparisons. The process of extracting and gathering meaningful features could be expanded further.

2. LITERATURE SURVEY

Prabu palanisamy et. al presented our system that they used for the SemEval-2013 Task 2 for doing Sentiment Analysis for Twitter data. They goted an F-score of 0.8004 on the test data set. They presented a lexicon-based method for Sentiment Analysis with Twitter data. They provided practical approaches to identifying and extracting sentiments from emoticons and hashtags. They also provided a method to convert non-grammatical words to grammatical words and normalize non-root to root words to extract sentiments. A lexicon-based approach is a simple, viable and practical approach to Sentiment Analysis of Twitter data without a need for training. A Lexicon based approach is as good as the lexicon it uses. To achieve better results, word sense disambiguation should be combined with the existing lexicon approach.Anna Jurek et. al we presented a new approach to lexicon-based sentiment analysis of Twitter messages. In the new approach, the sentiment is normalized, which allows us to obtain the intensity of sentiment rather than positive/negative decision. A new evidence-based combining function was developed to improve performance of the algorithm in the cases where a mixed sentiment occurs in a message. The evaluation was performed with the Stanford Twitter test set and IMDB data set. It was found from the results that the two new functions improve performance of the standard lexicon-based sentiment analysis algorithm. It could be noticed that the method is more appropriate for short messages such as tweets. When applied with long documents the method performed significantly better on the sentence than on the document level. Following this, this intention was to investigate the relationship between

the amount and the level of negative sentiment related to a public demonstration and the level of violence and disorder during the event. In other words, they aimed to ascertain if sentiment analysis could be applied as a supportive tool while predicting a level of disruption prior to public events. As a first step in this study, we decided to examine Twitter as a source of data. Four different demonstrations were selected, and the negative sentiment related to these events was analyzed over 6 days prior to each event.

3. EXISTING SYSTEM:

Following the case study and several analyses they were able to reveal that there was a relationship to some extent between the negative sentiment and the level of disorder during the EDL events. Further research is however required in this area to provide more accurate findings and conclusions. At the current stage we can, however, conjecture that sentiment analysis of social media content can provide valuable, security-related information regarding some upcoming public events. In the next step they wish to collect more data related to public events and further investigate the relationship between negative sentiment and the level of violence and disorder during events. Following this, we aim to develop a predictive model that can be used by police services as a single tool to help indicate violence propensity.

4. PROPOSED METHOD :-

Lexicon algorithm is used to determine the sentiment expressed by a textual content. This sentiment might be negative, neutral, or positive. It is possible to be sarcastic using only positive or neutral sentiment textual contents. Hence, lexicon algorithm can be useful but yet insufficient for sarcasm detection. It is necessary to extend the lexicon algorithm to come out with systems that would be proven efficient for sarcasm detection on neutral and positive sentiment textual contents. In this paper, two sarcasm analysis systems both obtained from the extension of the lexicon algorithm have been proposed for that sake. The first system consists of the combination of a lexicon algorithm and a pure sarcasm analysis algorithm. The second system consists of the combination of a lexicon algorithm and a sentiment prediction algorithm. In this work, the lexicon algorithm has been

Journal of Science and Technology
ISSN: 2456-5660 Volume 9, Issue 01 (Jan -2024)
www.jst.org.in *DOI:https://doi.org/10.46243/jst.2024.v9.i01.pp116-124*

extended in two ways to generate two systems that could be more efficient for sarcasm analysis, especially on neutral and positive sentiment textual contents.

5.SYSTEM ARCHITECTURE :

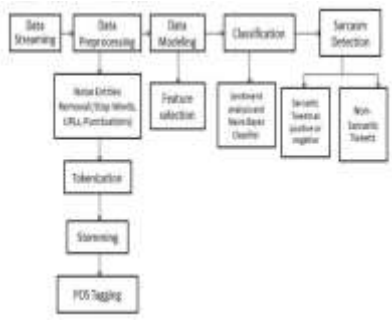

Figure no 1. System Architecture

6.RESULT :

Above screen descibes sarcasm result based on input dataset.

Above screen discribed sarcasm content like positive and nagetive.

7.CONCLUSION

The aim of this study was to propose ways to extend the lexicon algorithm to build systems that would be more efficient for sarcasm detection. This aim has been successfully met as two systems have been developed to address this situation. However, in the first system, it had been noticed that the training set of the sarcasm analysis algorithm must be relevant to the actual data that need to be analyzed to obtain meaningful results and to improve the accuracy of the system. The second system constitutes a vast area of study. Some work needs to be done to develop a system that would allow the collection of environmental details under which the textual contents would be made on social media platforms. A consolidated way of computing the sentiment polarity of the environments based on their details should also be developed.

8. REFERENCES

[1] Cambridge University Press, 2023. Sarcasm. [Online] Available At: Https://Dictionary.Cambridge.Org/Dictionary/English/Sarc Asm [Accessed 20 January 2018].

[2] Palanisamy, P., Yadav, V., & Elchuri, H. (2023). Serendio: Simple And Practical Lexicon- Based Approach To Sentiment. 543-548.

[3] Jurek, A., Mulvenna, M. D., & Bi, Y. (2021). Improved Lexicon-Based Sentiment Analysis For Social Media Analytics. Springeropen, 4-9.

[4] Kiilu, K. K., Okeyo, G., Rimiru, R., & Ogada, K. (2022). Using Naïve Bayes Algorithm In Detection Of Hate Tweets. International Journal Of Scientific And Research Publications, 99- 107.

[5] Rathan, K., & Suchithra, R. (2022). Sarcasm Detection Using Combinational. Imperial Journal Of Interdisciplinary Research, 546-551.

[6] Sathya, R., & Abraham, A. (2022). Comparison Of Supervised And Unsupervised. International Journal Of Advanced Research In Artificial Intelligence, 34-38.

[7] Dataquest, 2022. Top 10 Machine Learning Algorithms For Beginners. [Online] Available At: <Https://Www.Dataquest.Io/Blog/Top-10-Machine-Learning-Algorithms-For-Beginners/> [Accessed 15 September 2022].

[8] Haripriya, V., & Patil, D. P. (2022). A Survey Of Sarcasm Detection In Social Media. International Journal For Research In Applied Science & Engineering Technology, 1748-1753.

[9] Musto, C., Semeraro, G., & Polignano, M. (n.d.). A Comparison Of Lexicon-Based Approaches For Sentiment Analysis Of Microblog Posts.

[10] Saxena, R., 2023. How The Naive Bayes Classifier Works In Machine Learning. [Online] Available At: Http://Dataaspirant.Com/2023/02/06/Naivebayes-Classifier-Machine-Learning/ [Accessed 10 February 2023].

[11] Gandhi, R., 2022. Naive Bayes Classifier. [Online] Available At: Https://Towardsdatascience.Com/Naivebayes-Classifier-81d512f50a7c [Accessed 10 February 2023].

[12] Ray, S., 2023. 6 Easy Steps To Learn Naive Bayes Algorithm (With Codes In Python And R). [Online] Available At: Https://Www.Analyticsvidhya.Com/Blog/2023/09/Naivebayesexplained/ [Accessed 10 February 2023].

[13] Aggarwal, S., & Kaur, D. (2023). Naïve Bayes Classifier With Various Smoothing. International Journal Of Computer Trends And Technology, 873-876.

[14] Subba Reddy Borra, G. Jagadeeswar Reddy And E. Sreenivasa Reddy, "Fingerprint Image Compression Using Wave Atom Transform", International Journal Of Advanced Computing, Vol. 48, No. 1, 2015.

[15] Subba Reddyborra, G.Jagadeeswar Reddy, E.Sreenivasa Reddy, "Classification Of Fingerprint Images With The Aid Of Morphological Operation And Agnn Classifier", Applied Computing And Informatics, 2017.

[16] Subba Reddy Borra, Akshaya, B. Swathi, B. Sraveena, B. Satya Sahithi. (2023). Machine Learning Algorithms-Based Prediction Of Botnet Attack For Iot Devices. Turkish Journal Of Computer And Mathematics Education (Turcomat), 14(03), 65–78. Https://Doi.Org/10.17762/Turcomat.v14i03.13938.

[17] Dr. B. Subba Reddy, D. R. Amrutha Nayana, G. Sahaja, G. Shanmukha Priya. (2023). Artificial Intelligence Tool For Fake Account Detection From Online Social Networks. Turkish Journal Of Computer And Mathematics Education (Turcomat), 14(1), 243–254. Https://Doi.Org/10.17762/Turcomat.v14i1.13528.

[18] Subba Reddy Borra, B Gayathri, B Rekha, B Akshitha, B. Hafeeza. (2023). K-Nearest Neighbour Classifier For Url-Based Phishing Detection Mechanism. Turkish Journal Of Computer And Mathematics Education (Turcomat), 14(03), 34–40. Https://Doi.Org/10.17762/Turcomat.v14i03.13935.

[19] Dr. Subba Reddy Borra, K.Harshitha, Kudithi Neha, K. Sindhusha, K. Akshaya. (2023). Block Chain Based Agriculture Crop Delivery Platform For Enforcing Transparency. Turkish Journal Of Computer And Mathematics Education (Turcomat), 14(2), 987–997. Https://Doi.Org/10.17762/Turcomat.v14i2.13925.

[20] Dr. B. Subba Reddy, S. Shresta, S. Sathhvika, P. Lakshmi Manasa Shreya. (2023). Detection Of Electricity Theft Cyber-Attacks In Renewable Distributed Generation For Future Iot-Based Smart Electric Meters. Turkish Journal Of Computer And Mathematics Education (Turcomat), 14(2), 64–77. Https://Doi.Org/10.17762/Turcomat.v14i2.13523.

[21] Dr. Subba Reddy Borra, K.Harshitha, Kudithi Neha, K. Sindhusha, K. Akshaya. (2023). Block Chain Based Agriculture Crop Delivery Platform For Enforcing Transparency. Turkish Journal Of Computer And Mathematics Education (Turcomat), 14(2), 987–997. Https://Doi.Org/10.17762/Turcomat.v14i2.13925.

[22] Dr. B. Subba Reddy, S. Shresta, S. Sathhvika, P. Lakshmi Manasa Shreya. (2022). Role Of Machine Learning In Education: Performance Tracking And Prediction Of

Students. Turkish Journal Of Computer And Mathematics Education (Turcomat), 13(03), 854–862. Https://Doi.Org/10.17762/Turcomat.v13i03.13175.

[23] B. Subba Reddy, V. Bhargavi., S. Samhitha, Y. Anjana, V. Saivaishnavi. (2023). Nlp-Based Supervised Learning Algorithm For Cyber Insurance Policy Pattern Prediction. Turkish Journal Of Computer And Mathematics Education (Turcomat), 14(03), 90–99. Https://Doi.Org/10.17762/Turcomat.v14i03.13941.

[24] Dr. B. Subba Reddy, P. Sai Hamshika, S. Aishwarya, V. Ashritha. (2022). Block Chain For Financial Application Using Iot. Turkish Journal Of Computer And Mathematics Education (Turcomat), 13(03), 844–853. Https://Doi.Org/10.17762/Turcomat.v13i03.13174.

AUTHORS :

First Author: Dr. SUBBA REDDY BORRA, Professor and Head,Department of Information Technology at Malla Reddy Engineering College for Women(UGC-Autonomous) in Dhulapally,Hyderabad.Email:bvsr79@gmail.com.

Second Author: T. KRUTHIKA is pursuing her B.Tech degree in Information Technology from MALLAREDDY ENGINEERING COLLEGE FOR WOMEN(AUTONOMOUS).

Third Author: T.SAI POOJITHA is pursuing her B.Tech degree in Information Technology from,Malla reddy engineering college for women(Autonomous).

Fourth Author: P. DEEPIKA is pursuing her B.Tech degree in Information Technology from Mallareddy Engineering College For Women(Autonomous).