Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 06, June-2019

Survey of Machine Learning for Sarcasm Detection

Payas Relekar
Dept. of Electronics and Telecommunication
Sinhgad College of Engineeering, Vadgaon (Bk)
Pune, India

Prof. Samadhan D. Mali
Dept. of Electronics and Telecommunication
Sinhgad College of Engineeering, Vadgaon (Bk)
Pune, India

***Abstract*—Sarcasm is a very unique feature of human expression. It is heavily context dependent, and most humans understand it by means of tone and other visual cues in spoken language. This is particularly difficult in online discussions where these cues are not present, and even humans may fail to understand a post as being sarcastic. Renewed interest in using Artificial Intelligence for natural language processing, and increasing communication via text has many researchers intrigued by this problem. In this paper, some of the recent research on sarcasm detection are studied and compared.**

***Index Terms*— Sarcasm, machine learning, reddit, twitter, communication, language**

I. INTRODUCTION

Sarcasm is a sophisticated form of communication in which the speaker conveys exact opposite of their intent. It is often associated with irony or satire due to this, as most popular usage of sarcasm is to mock, insult or amuse. Its use is even associated with workplace anger[1] and language development in children[2]. However, it is not possible to provide visual and tonal cues in online discussions.

Some large online communities have adopted a specific annotation to express sarcasm in case readers fail to understand it on their own. Twitter users use '#sarcasm' or '#not' at the end of a tweet to indicate a sarcastic post. Reddit users on the other hand annotate with '/s' to self-declare their post as sarcastic instead. Both of these annotations provide what online communication inherently lacks to detect irony: linguistic cues. However, sarcastic posts sarcastic posts are not guarranteed to contain such labels at all. Hence, detection of sarcasm in such posts again relies on limited information: associated context.

In recent years, detecting sarcasm in online discussions has garnered some interest, and not just for research[3]. Since simple word association fails, more sophisticated ways including Machine Learning are being used. In this paper, Some the findings of these studies and their methodology are explored in this paper. In addition, this paper focuses on following questions in particular:

- What are the commonly used datasets and features?
- Which algorithms are popular and why?
- How can these algorithms or datasets be improved upon?

Our findings are presented below, and hope they'll aid further research.

II. PREREQUISITES

A. Defining sarcasm

According to *Oxford Dictionary*, irony is "the use of words to express something other than and especially the opposite of the literal meaning of a sentence", while *Wikipedia* defines Sarcasm as "a sharp, bitter, or cutting expression or remark; a bitter gibe or taunt". As such sarcasm can be thought of as an instance of irony that involves an emotional aspect, possibly one of the aggressions.

B. Datasets

Although research on this topic is fairly recent, Twitter is single most popular source of data for written language studies so far. And there are good reasons for that. Let's borrow a famous term given to big data, volume, variety and velocity (3V). Volume, defined as the quantity of generated and stored data. According to Twitter own website, 313 million people took access every month. Assuming every user tweets twice a day that makes 616 million tweets only ina month. Variety, The type and nature of the data. Tweet made by many people has different characteristic depending on many aspects, like geographical, political or current trending. Gender and age factor also play a big factor in tweet varied. This makes Twitter a potential platform to be analyzed in many areas of studies. Velocity, in this context, the speed at which the data is generated and processed to meet the demands and challenges that lie on the path of growth and development [4].

Two of the features that make twitter attractive are hashtags and retweets. Hashtags are labels that a user applies to their own post as a string of characters preceeded by hash or pound sign(#). This allows classifying tweets much easier. Retweets are reposts of a tweet which show its relative popularity.

However, until very recently twitter had an arbitrary limit of 160 characters per post. This results in distinct grammar, which means applying models trained on tweets to other data isn't particulary good idea. Thankfully, there are now some more datasets available, notably Amazon and Reddit.

Amazon dataset includes reviews such as product ratings, text and helpfulness votes. Product metadata such as product description, category information, product price, product brand and image features.

And for the newest dataset SARC, short for Self Annotated Reddit Corpus[5], utilises Reddit comments. It was originally created in 2017, and updated in 2018. This is by far the best dataset in our observation, with multiple advantages. It is an order of magnitude larger than other datasets, and includes detailed subsets for politics as well as balanced and unbalanced sets for distinctive training. Also, reddit comment limit is of 40000 characters, allowing much longer posts and normal human language.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 06, June-2019

Table I
Comparison of Popular Datasets

Corpus	Dataset	Sarcastic	Total
IAC	Joshi et al. (2015)[6]	751	1502
	Oraby et al. (2016)[7]	4.7K	9.4K
Twitter	Joshi et al. (2016) [8]	4.2K	5.2K
	Bamman & Smith (2015)[9]	9.7K	19.5K
	Reyes et al. (2013)[10]	10K	40K
	Riloff et al. (2013)[11]	35K	175K
	Ptacek et al. (2013)[12]	130K	780K
Reddit	Wallace et al. (2015)[13]	753	14124
	SARC (2017)	1.34M	533M

Reddit is newly popular forum and has threaded comments, which means full context of the post is preserved and can be utilized. Due to all these reasons, this dataset is becoming increasing popular with newer researches on this topic.

III. LITERATURE SURVEY

Sarcasm detection using machine learning is relatively recent field to have garnered interest. Most of the previous work can broadly classified in two major categories, based on their architecture and methodology.

1) Content-based Models: These models are relatively simple and straightforward. They try to analyse and classify lexical and grammatic structure of text to identify sarcasm. Tepper man et al. (2006) [14] analysed spoken dialogue systems for sporadic and spectral cues. Carvalho et al. (2009) [15] tried to use emoticons, quotation marks and other linguistic features like interjections, positive predicates and gestural clues to weigh various elements of statements. Gonzalez-Ib´ anez´ et al. (2011) [16] also used emoticons for weighing sarcasm possibility, but using tweets. Davidov et al. (2010) [17], Tsur et al. (2010) [18] constructed classifiers by identifying patterns in language syntax. Riloff et al. used the priciple that sarcasm is stated by using positive sentiment words to communicate negative scenarios. Joshi et al. (2015) used combination of multiple lexical features along with pragmatics, implicit and explicit context clues. For explicit cases, they used they used relevant features to find inconsistently stopped sentimental descriptions. For implicit scenarios, they extended Riloff et al. (2013) by identifying verb-noun phrases containing contrast in positive and negative polarities. Since these models only rely on what is in the statements and for the most part ignore contextual clues, their accuracy hits diminishing returns for more complex analysis. Sarcasm or irony being almost exclusively context dependent, it is insufficient to do using only lexical, grammatical and syntactic clues. However, these research findings were important in that they paved the way forward by making clear that content-based models are not sufficient on their own.

2) Context-based Models: Since internet as become widespread and reliable means of communication, more and more people are using it for regular discussions. Combined with relative/pseudo anonymity that internet provides, usage of sarcasm on online platforms has tended to increase in recent years. However, a lot of these posts, especially on microblogs or forums and social media are highly affected by grammatical mistakes by people with English as their non-native language or incomplete support for their native language. These posts also contain information that's highly temporal and context dependent. Carvalho et al. (2009), Wallace et al. (2014) showed that lexical clues alone are not sufficient and fail in situations where humans require and can use of additional context. They also stated importance of topical information and identity of the speaker/commenter to be associated to add such context to a text. Poria et al. (2016) [19] use emotional, sentiment and personality representation of the input text as such additional information. Redesigning et al. (2015) [20], Zhang et al. (2016) [21] used historical posts of users to determine sarcastic tendency of individual poster. Khatri et al. (2015) [22] explored user's histories to identify their historic opinion on the subject for contrasting statements. Wallace et al. (2015) used nouns and phrases to find general sentiment on the topic in the forum and built context specific to the forum. Such modelling shows how community hive mind works and opinions form by being part of one. It also provides additional context compared to history of individual person alone. ask.

Pennington et al. (2014) [23] devised a novel technique of converting words to vector matrices. Such vectors can be used over bigger text to generate embeddings, which are essentially text converted to fewer dimensions represented in numbers. Amir et al. (2016) [24] created user models by generating embeddings that capture homophily using a technique similar to that of Milokov et al. (2014) [25]. Such embeddings convert text into word vectors of higher dimensions. Such dimensions allow finding similarities between different words as well as relations between them trivial via simple functions such as sigmoid functions.

IV. MACHINE LEARNING ALGORITHMS

A. Classification Algorithm

Various classification algorithms are organised by their approaches. First approach is machine-learning and second is rule-based. In machine-learning, a model to classify, arrange or predict data through statistical process is formed. While in rule-based approach, semantic, syntactic, and stylistic properties of the text, such as structural and lexical attributes are analysed to classify and predict sentiment of the post.

B. Supervised Learning

Supervised learning is comparatively simpler kind of machine learning and often can construct a decently accurate function from labelled dataset [26]. This is possible because for during supervised learning, expected results of the model are already provided in the training dataset. This allows the model to pre-emptively train and optimize the cost function to accurately align with actual results. Supervised learning can further be used as base for other algorithms, such as naive Bayes, decision tree, logistic regression, etc. Support vector machine (SVM) also applied as the preferred algorithm for sarcasm sentiment analysis.

C. Semi-supervised

Semi-supervised algorithms, sometimes also called inductive learning or transductive learning combines aspect of Supervised and unsupervised learning methods. Typically, this means using small labelled dataset in tandem with large unlabelled data.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 06, June-2019

D. Structured Learning

Structured learning tries to generalize standard paradigms of supervised learning, classification, and regression. In other words, it relies on finding a function that minimizes some loss over a training set. It is particularly useful when available training data is very large or available features in dataset are limited.

E. Hybrid Approach

Hybrid learning or transfer learning are very modern approaches and becoming popular quickly. This is combining one or more pretrained models. These models may have used entirely different datasets for training, and hence need to be carefully evaluated and applied. However, considering extreme requirements for training data as well as training time and resultant model size, it may be efficient to take this approach over starting from scratch.

F. Neural Network

Neural networks are most common machine learning algorithms because their similarity with human brain structure and function. They are among fundamental concepts of Artificial Intelligence and machine learning and widely understood. They also have variety of implementations including Convoluted Neural Network, Recurrent Neural Network etc. These allow flexible structure and quick feedback and thus are favoured as baseline models for comparison.

V. GENERAL FINDINGS

For all the studies we've seen so far, it is clear that the field is still pretty recent and wildly different methods showing no sign of convergence or best practises. However, what is becoming increasingly clear, is that simple neural networks may not be sufficient and higher availability and quality of context leads to better results. Moreover, lexical and structural features, when considered in isolation of each other, may even lower overall accuracy of the results.

A. Issues in Sarcasm Detection

1) Issues with Datasets: Most glaring issue with current popular datasets such as Twitter and Reddit are that these are anonymous networks, and as such lack social hierarchy in real world. This leads to unusually high amount of sarcasm in certain topics, such as sports or politics. While this may lead to larger dataset, they may not be accurately match real world usage of sarcasm.

Another aspect with using hashtags such as #sarcasm, #not or Reddit's /s tag is that they may be absent for some sarcastic posts or even used incorrectly by unsure users. While dataset creators for e.g. SARC have taken proper precautions to avoid such scenarios, with automated collection of comments these datasets utilise, it may not be possible to minimize impact of such scenarios.

2) Issues with Features: Using contextual information such as parent comment or subforum/topic of discussion as added features during training increases overall accuracy. Therefore, more such features should be explored, for example, user history, subforum/topic comment history and profiling of these individual features etc. may prove good future areas to research.

VI. CONCLUSION AND FUTURE WORK

In this paper, commonly used approaches for sarcasm detection using machine learning are. In this paper, both rule-based approaches as well as machine learning algorithms are discussed. But, because of different nature, shape and application of sarcasm existing in real life, this problem is too wide and challenging to be made as a generalized formula. Future work regarding sarcasm detection can be concluded as follows:

- *Datasets:* Future datasets should include more contextual features temporal reference as well as more immediate information on topic.
- *Author profiles:* Profiling authors and subforums through their history should lead to highly effective feature.

All in all, newer studies are showing higher accuracy in results as they're already beginning to apply some of the suggestions. We look forward to observe future research in this area.

REFERENCES

[1] K. R. Calabrese, "Interpersonal Conflict and Sarcasm in the Workplace ", Genetic Social and General Psychology Monographs, pp. 459-494, 2000.
[2] S. McDonald, "Exploring the Process of Inference Generation in Sarcasm:," Brain and Language, Volume 68, no. 3, pp. 486-506, 1999.
[3] BBC, "US Secret Service seeks Twitter sarcasm detector,"https://www.bbc.com/news/technology-27711109", 2014.
[4] M. Hilbert, "Big Data for Development: A Review of Promises and Challenges," Development Policy Review, Volume 1, no. 34, p. 135–174, 2016 .
[5] Mikhail Khodak, Nikunj Saunshi, and Kiran Vodrahalli, "A large selfannotated corpus for sarcasm", arXiv preprint arXiv:1704.05579, 2017.
[6] Joshi, A., Sharma, V., and Bhattacharyya, P., "Harnessing context incongruity for sarcasm detection. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics", pages 757762. Association for Computational Linguistics, 2015.
[7] Oraby, S., Harrison, V., Reed, L., Hernandez, E., Riloff, E., and Walker, M., "Creating and characterizing a diverse corpus of sarcasm in dialogue", In Proceedings of the SIGDIAL 2016 Conference, pages 31–41. Association for Computational Linguistics, 2017.
[8] Joshi, A., Bhattacharyya, P., and Carman, M. J., "Automatic sarcasm detection: A survey", arXiv, 2016.
[9] Bamman, D. and Smith, N. A., "Contextualized sarcasm detection on twitter", Association for the Advancement of Artificial Intelligence, 2015.
[10] Reyes, A., Rosso, P., and Veale, T., "A multidimensional approach for detecting irony in twitter", Data Knowledge Engineering, 47(1), 2013.
[11] Riloff, E., Qadir, A., Surve, P., Silva, L. D., Gilbert, N., and Huang, R., "Sarcasm as contrast between a positive sentiment and negative situation", In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pages 704–741. Association for Computational Linguistics, 2013.
[12] Ptacek, T., Habernal, I., and Hong, J., "Sarcasm detection on czech and english twitter," 25th International Conference on Computational Linguistics: Technical Papers, pages 213-223, 2014.
[13] Wallace, B. C., Choe, D. K., and Charniak, E. Sparse, "Contextually informed models for irony detection: Exploiting user communities, entities, and sentiment", In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics, pages 1035-1044. Association for Computational Linguistics, 2015.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 06, June-2019

[14] Joseph Tepperman, David Traum, and Shrikanth Narayanan, "yeah right: Sarcasm recognition for spoken dialogue systems", In Ninth International Conference on Spoken Language Processing, 2006.

[15] Paula Carvalho, Luıis Sarmento, Mario J Silva, and Eugenio De Oliveira,´ "Clues for detecting irony in user-generated contents: oh...!! it's so easy", In Proceedings of the 1st international CIKM workshop on Topicsentiment analysis for mass opinion, pages 53-56. ACM, 2009.

[16] Roberto Gonzalez-Ibanez, Smaranda Muresan, and Nina Wacholder, "Identifying sarcasm in twitter: a closer look", In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers-Volume 2, pages 581-586.
Association for Computational Linguistics, 2011.

[17] Dmitry Davidov, Oren Tsur, and Ari Rappoport, "Semi-supervised recognition of sarcastic sentences in twitter and amazon", In Proceedings of the fourteenth conference on computational natural language learning, pages 107–116. Association for Computational Linguistics, 2010.

[18] Oren Tsur, Dmitry Davidov, and Ari Rappoport, "Icwsm-a great catchy name: Semi-supervised recognition of sarcastic sentences in online product reviews", In ICWSM, pages 162-169, 2010.

[19] Soujanya Poria, Erik Cambria, Devamanyu Hazarika, and Prateek Vij, "A deeper look into sarcastic tweets using deep convolutional neural networks", In Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, pages 16011612, 2016.

[20] Ashwin Rajadesingan, Reza Zafarani, and Huan Liu, "Sarcasm detection on twitter: A behavioral modeling approach", In Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, pages 97–106. ACM, 2015.

[21] Meishan Zhang, Yue Zhang, and Guohong Fu, "Tweet sarcasm detection using deep neural network", In Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, pages 2449–2460, 2016.

[22] Anupam Khattri, Aditya Joshi, Pushpak Bhattacharyya, and Mark Carman, "Your sentiment precedes you: Using an author's historical tweets to predict sarcasm", In Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pages 25-30, 2015.

[23] Jeffrey Pennington, Richard Socher, Christopher D. Manning, "GloVe: Global Vectors for Word Representation", Empirical Methods in Natural Language Processing (EMNLP), pages 1532-1543 2014.

[24] Silvio Amir, Byron C Wallace, Hao Lyu, and Paula Carvalho Mario J´ Silva, "Modelling context with user embeddings for sarcasm detection in social media", arXiv preprint arXiv:1607.00976, 2016.

[25] Quoc V Le and Tomas Mikolov, "Distributed representations of sentences and documents", arXiv preprint arXiv:1405.4053, 2014.

[26] M. Mohri, A. Rostamizadeh ve A. Talwalkar, "Foundations of Machine Learning", London: The MIT Press, 2012.