A Research on Detection of Sarcasm using Machine Learning Techniques

Amruta K. Chimote, S.R. Tandan

Abstract— Today in the era of flooding information on online forums, social media where decision making of customer(user) is assisted in all possible ways. Almost all types of applications have their own assistant or chatbots to guide user for his/her query. Use of assistants and chatbots gives real experience and at the same time for admin it is performing the role of customer care executive. To make it more realistic most of assistants and chatbots are built in conversational format which guide user to get correct information. It is necessary to have knowledge of user sentiments while assistance is provided. Sarcasm is one of complex human sentiment which is used to convey disagreement using positive words. This paper provides review on various approaches used to detect sarcasm present in textual data using supervised and unsupervised approaches of machine learning. It is observed many authors used neural networks algorithms to detect sarcasm. Efficiency and accuracy of detection can be increased with combination of different approaches.

I. INTRODUCTION

1.1 What is Sarcasm? (Some Definition)

Sarcasm is one of the sentiments representing something positive with negative intent. Sarcasm is defined by so many people with different approach. Some of definitions are :

As said by Mondher Bouazizi Sarcasm is a sophisticated form of irony widely used in social networks and micro blogging websites.

As stated in Cambridge Dictionary by Francesca sarcasm is considered as wit but with high intelligence.

1.2 Sentiment analysis

Sentiment analysis is growing topic in the field of review mining and needed for human machine interaction as well while training any machine regarding human emotions and sentiments. Detection of various sentiments from text/speech will lead to correct decision making. (AshrafUllah, September 2014)

Above is the wheel of sentiments which shows the various sentiments of human, belongs to which category of mood. Sarcasm is one of the sentiments used by people to express their views about people, place, product or any situation over social media. Detection of sarcasm present in any speech/text will be helpful to retrieve correct meaning of sentence, user review and any speech/text available on social media which may influence people to make any decision.

1.3 What is Sarcasm detection?

Identifying sarcasm present in given data of text/speech which may include instructions, blogs, views or reviews is known as sarcasm detection. Sarcasm analysis have become emerging topics of research in opinion mining and sentiment analysis in recent years but most of the work is focused on data in the English language. Detection of sarcasm will further lead to identification of psychology, mood and sometimes health condition of people, which helps in making human machine interaction easier and in some cases it will make diagnosis of brain injuries at early stages. (Katherine P. Rankin et. al 2010 and Mcdonald, Skye et. al 2005). Continuous observation on sarcastic reviews or comments can also help to prevent suicidal cases in youth (Aditya Joshi, et.al.).

Revised Version Manuscript Received on 16 September, 2019.
 Amruta K. Chimote1, Research Scholar, Department of Computer Science & Engineering, Dr. C.V.Raman University, Bilaspur(CG), India..
 S.R. Tandan, Associate professor , Department of Computer Science & Engineering, Dr. C.V.Raman University, Bilaspur(CG), India.

II. REVIEW OF LITERATURE

Wide range of applications are there which based on sentiments, it's obvious so many people get attracted towards this area. As sarcasm is the complex sentiment to identify various researchers contributed in detection of sarcasm in present text. Most of them have focused on social media text for analysis. Majorly we found that most of work done by using machine learning techniques. Researchers used classification, neural network and combinational approach for this.

Following papers discussed about the classification methods for sarcasm detection.

In the paper title itself sarcastically said that "Detecting Sarcasm is extremely easy " where researchers analysed the performance of system on datasets from Twitter comments and Amazon product reviews. This analysis helped for recommendations for this Methodology Used is Classification Algorithm (Nielsen, June 5, 2018)

Multinomial Naïve Bayes (MNNB)some of the researchers like (P, April 2018) used multinomial naïve bayes (MNNB) and support vector machines approaches of machine learning to detect sarcasm. They used tweets to detect sarcasm.

(Wiken, August 2015) prepared model on detection of irony and sarcasm which classify text into sarcastic and non-sarcastic groups they also used supervised machine learning techniques on amazon and twitter reviews.

It is observed that for better results, neural networks played major role, following papers focused on the use of neural network algorithms such as CNN, LSTM and RNN, etc

(Foster, October 31, 2018) manually annoted a freely available English sentences, where polarity of expressions is classified. To check polarity CNN and LSTM was used by researchers. Series of experiments were carried out for this detection.

In the Paper , "Sarcasm SIGN: Interpreting Sarcasm with Sentiment Based Monolingual Machine Translation " ,authors have interpreted sarcastic sentence into non sarcastic way, focus was given on generation of original sarcasm text into non sarcastic text for this machine translation and RNN used. (Reichart, July 30 - August 4, 2017.)

"Encoding Sentiment Information into Word Vectors for Sentiment Analysis" , This paper focused on following points: (1) feed forward neural network was used to encode lexicons into word vectors .(2) During the training of classifier like CNN ,The word vectors are fine-tuned for supervised training of data and the sentiment lexicon . (Zhe Ye, August 20-26, 2018)

In Contextual based detection of sarcasm some researchers like (Veale, September 7–11, 2017) accurately detect sarcasm based on speaker's mood. This work was carried over tweets leading to new post which is useful cue for detection of sarcasm. Methodology Used are Combination of CNN/LSTM (long short-term memory RNN) .

Sarcasm detection can be made easy if context of statements will be analysed to get correct discourse of the sentences. This will help in auto question answering systems, efficient chatbot and Human Machine Interaction. Following papers speaks about the context driven approach.

In a Contextual Sarcasm Detector (CASCADE), which works on hybrid approach of both context- and content-driven modelling for sarcasm detection in online forums. Further it leads to discourse analysis of sentences. Methodology Used: CNN (Devamanyu Hazarika, August 20-26, 2018)

In Lexicons based sentiment analysis. (Slonim, August 20-26, 2018.) have proposed novel approach to get composition of sentiment from large and not labelled corpus. Authors used unigrams and bigrams on set of lexicons to detect variety of sentiment composition.

Social media is continuously attracting researchers for sarcasm detection, authors explored two methods for representing the context of textual sarcasm detection, Bayesian approach that directly represents authors' natural tendency to be sarcastic, and strong embedding approach that can learn relativity between the author and the text. Methodology Used: RNN with GRU cells. (Y. Alex Kolchinski, October 31 - November 4, 2018)

"Introduction to the Special Issue on Language in Social Media: Exploiting Discourse and Other Contextual Information" ,

In this paper authors simply applying

Some of text mining tools are take into consideration. It focuses on spoken as well as written text. This helps in understanding the more correct meaning of the processed data on social media which further leads to discourse analysis. This helps to understand context from linguistic and computational linguistic point of view. Various NLP approaches are also used on social media text for contextual information. (Farah Benamara)

Previous paper focused on context of text or spoken data, Authors have discovered the context from conversation. Overall concept of conversation is understood using LSTM networks. Context is analysed over social media. Methodology Used: Qualitative analysis of the attention weights produced by the LSTM models (Debanjan Ghosh)

"Reasoning with Sarcasm by Reading In-between", Authors revisit the notion of modelling contrast in order to reason with sarcasm. More specifically, authors propose an attention-based neural model that looks in between instead of across, enabling it to explicitly model contrast and incongruity. Authors conducted extensive experiments on six benchmark datasets from Twitter, Reddit and the Internet Argument Corpus. The proposed model not only achieves state of-the-art performance on all datasets but also enjoys improved interpretability.

Methodology Used:

Multi-dimensional Intra-Attention Recurrent Network (MIARN) (Yi Tayy, July 15 - 20, 2018)

(Taradheesh Bali, December 12 2016) proposed a novel approach to detect sarcasm on social media domain like twitter where context of discussion is understood between two people. This focuses on context discussion between audience and generator of sarcastic views. Lexical and pragmatic approaches are used for this model.

969

Combinational approach is also used for rhetorical questions, multi-view sentiments corpus data which used SVM and LSTM together for better results by

(Shereen Oraby, 15-17 August 2017)

(MessinaMilan, April 3-7, 2017) respectively.

Some of medical science background based papers speak on the detection of brain injuries based on spoken language of surveyed people. Detection of sarcasm can also help medical science to detect injuries, and linguistic development (autism as well) by (Rambam, 2005) (Channon S) (Katherine P. Rankin, 2009 Oct 1)

III. SCOPE OF WORK & RESULTS

Scope of natural language processing is wide enough which leads to Machine Translation, Text Simplification, Spam Filter, Auto-Predict, Question Answering, Natural Language Generation and sentiment Analysis. According to wheel of sentiments, sarcasm belongs to anger segment and that too again classified into critical expression. So, we proposed model which will classify data based on its expression and will precisely classify into sub categories of anger which will lead to correct detection of sarcasm into provided speech/text. Annotation regarding each sub category of anger sentiment will be provided and it will be checked against the model developed.
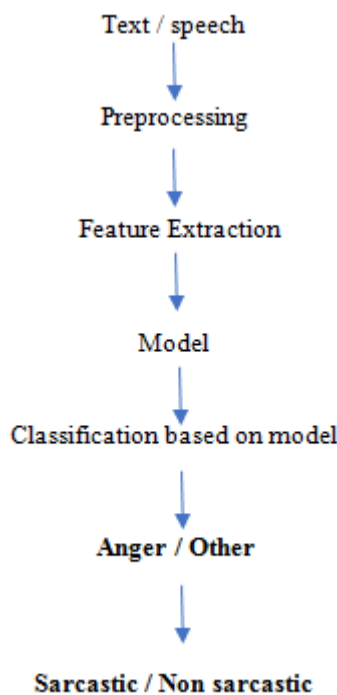
IV. CONCLUSION

According to above reviewed work of various authors, it is observed detection of sarcasm a crucial work in linguistic. This will help in serving customer reviews, online discussion forums, chatbots and even detection of brain injuries based on ability of person's writing skill as well.

REFERENCES

1. AshrafUllah, K. B. (September 2014). Mining opinion components from unstructured reviews: A review. Journal of King Saud University - Computer and Information Sciences, 26, Pages 258-275.
2. Channon S, P. A. (n.d.). Social cognition after head injury: sarcasm and theory of mind.
3. Debanjan Ghosh, A. R. (n.d.). Sarcasm Analysis Using Conversation Context. Creative Commons Attribution-Non Commercial-No Derivatives 4.0 International (CC BY-NC-ND 4.0) license Computational Linguistics Volume 44, Number 4.
4. Devamanyu Hazarika, S. P. (August 20-26, 2018). CASCADE: Contextual Sarcasm Detection in Online Discussion Forums . 27th International Conference on Computational Linguistics (pp. 1837–1848). New Mexico, USA: Computational Linguistics.
5. Farah Benamara, P. I. (n.d.). Introduction to the Special Issue on Languagein Social Media: Exploiting Discourse andOther Contextual Information . Creative Commons Attribution-Non Commercial-No Derivatives 4.0 International (CC BY-NC-ND 4.0) license Computational Linguistics Volume 44, Number 4 .
6. Foster, R. K. (October 31, 2018). Sentiment Expression Boundaries in Sentiment Polarity Classification . 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, (pp. 156–166). Brussels, Belgium,.
7. Katherine P. Rankin, A. S.-T. (2009 Oct 1). Detecting Sarcasm from Paralinguistic Cues: Anatomic and Cognitive Correlates in Neurodegenerative Disease . Published in final edited form as: Neuro image(doi: 10.1016/j.neuroimage.2009.05.077).
8. MessinaMilan, D. N. (April 3-7, 2017). A Multi-View Sentiment Corpus . 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1 (pp. 273–280). Valencia, Spain: Computational Linguistics.
9. Nielsen, N. P. (June 5, 2018). Detecting Sarcasm is extremely easy ;-) . Proceedings of the Workshop on Computational Semantics beyond Events and Roles (SemBEaR-2018) (pp. 21–26). New Orleans, Louisiana: SemBEaR-2018.
10. P, R. K. (April 2018). Sarcasm Detection Using Machine Learning Techniques. International Publication of Recent Scientific Research, 26368-26372.
11. Rambam, S. G.-T. (2005). The Neuro anatomical Basis of Understanding Sarcasm and Its Relationship to Social Cognition . Medical Center Neuropsychology Copyright 2005 by the American Psychological Association 2005, Vol. 19, No. 3, , 288 –300 .
12. Reichart, L. P. (July 30 - August 4, 2017.). Sarcasm SIGN: Interpreting Sarcasm with Sentiment Based Monolingual Machine Translation . Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (pp. 1690–1700). Vancouver, Canada: Computational Linguistics.
13. Shereen Oraby, V. H. (15-17 August 2017). Are you serious?: Rhetorical Questions and Sarcasmin Social Media Dialog . Proceedings of the SIGDIAL 2017 Conference (pp. 310–319). Saarbrucken, Germany: SIGDIAL.
14. Slonim, O. T.-R.-H. (August 20-26, 2018.). Learning Sentiment Composition from Sentiment Lexicons . 27th International Conference on Computational Linguistics (pp. 2230–2241). Santa Fe, New Mexico, USA: Computational Linguistics.

15. Taradheesh Bali, N. S. (December 12 2016). Sarcasm Detection Building a contextual hierarchy . Proceedings of the Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media , , (pp. 119–127). Osaka, Japan.

16. Veale, A. G. (September 7–11, 2017). Magnets for Sarcasm: Making Sarcasm Detection Timely, Contextual and Very Personal. 2017 Conference on Empirical Methods in Natural Language Processing, . (pp. 482–491). Copenhagen, Denmark: Empirical Methods in Natural Language Processing,.

17. Wiken, E. F. (August 2015). Automatic irony- and sarcasm detection in Social media. UPTEC F15 045 .

18. Y. Alex Kolchinski, C. P. (October 31 - November 4, 2018). Representing Social Media Users for Sarcasm Detection . Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, (pp. 1115–1121). Brussels, Belgium.

19. Yi Tayy, L. A. (July 15 - 20, 2018). Reasoning with Sarcasm by Reading In-between. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Long Papers) (pp. 1010–1020). Melbourne, Australia: Computational Linguistics.

20. Zhe Ye, F. L. (August 20-26, 2018). Encoding Sentiment Information into Word Vectors for Sentiment Analysis . Proceedings of the 27th International Conference on Computational Linguistics (pp. 997–1007). New Mexico, USA: Computational Linguistics.