International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-4, November 2019

Automatic Sarcasm Detection with Textual and Acoustic Data

Steve Michael, Amalia Zahra

Abstract: *This paper takes focus on the area of automatic sarcasm detection. Automatic sarcasm detection is crucial due to the needs of sentimental analysis. The rapid development of automatic speech recognition and text mining and the large amount of voice and text data opens a broader way for researchers to open new method and improves the accuracy of automatic sarcasm detection. We observe approaches that have been used to detect sarcasm, kind of data and its features including the rises of context to improve the accuracy of automatic sarcasm detection. We found that some context cannot be reliable without the presence of other context and some approaches are very dependent on the dataset. Twitter is being used by researchers as the main mine for sentimental analysis, we notice that at some aspect it still has a flaw because it is dependent to some Twitter's special feature that will not be found in other usual text data like hashtags and author history. Besides that, we see that the small amount of research about automatic sarcasm detection through acoustic data and its correlation with textual data could make a new opportunity in the area of sarcasm detection in speech. From acoustic data, we could get both acoustic features and textual features. Sarcasm detection with voice has the potential to get higher accuracy since it can be extracted into two data types. By describing each beneficial method, this paper could be a brief way to sarcasm detection through acoustic and textual data.*

Keywords**:** *textual sarcasm detection, acoustic sarcasm detection, contextual features*

I. INTRODUCTION

The research of speech recognition is not a new hot thing anymore in the computer science field. Even though this topic has been discussed for several years, it never stops to sharpen its functionality in helping humanity through the technology of using speech in daily life. Automatic sarcasm detection may be classified as a mature topic in speech recognition and text mining area. Sarcasm is a form of figurative language where the literal meaning of words does not hold, and instead, the opposite interpretation is intended. Sarcasm is closely related to irony—in fact, it is a form of irony. People often use different tonal model and certain gestural clues to express their sarcasm while speaking. In the textual data, the absence of tonal and gestural clues makes sarcasm detection more challenging [1].

Manuscript published on November 30, 2019.
***** Correspondence Author
Steve Michael*, Computer Science Department, BINUS Graduate Program - Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480.
Amalia Zahra, Computer Science Department, BINUS Graduate Program - Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license http://creativecommons.org/licenses/by-nc-nd/4.0/

Since the first research of sarcasm detection in speech, the number of studies in sarcasm detection keeps growing. By looking deeper into the root, the main aim of detecting sarcasm is to promote the accuracy of the sentimental analysis. Sarcasm plays a big role in sentiment analysis, especially in some task such as review summarization in an NLP application [2], recommendation systems, etc. Sarcasm is crucial to sentiment analysis because it delivers an implied opposite meaning of a sentence in a conversation. It allows human to express negative connotation using irony in communication. The words "Yeah, I love Math" could refer to two different meanings which are it is rather the person loves Math or the opposite. There have been a lot of challenges that have been overcome through research, and there is some still untouched. Sarcasm is multimodal, it builds/involves tones, body-language, linguistic artifacts [2].

Based on the number of previous studies so far, most research is still focused through textual features, compared to studies using sound features. In terms of datasets, it can also be seen clearly that textual data that can be processed is certainly more in number than sound. Thus, there is still not much research in this theme conducted in acoustic field.

II. DATASETS, APPROACHES AND PARAMETERS

Several aspects determine how the process of sarcasm detection will be carried out and the level of accuracy obtained. This is because related aspects such as datasets, approaches, and features used in previous studies tend to vary so that the performance of each study cannot be compared directly.

A. Datasets

Voice Data. In "Yeah, Sure I did the right thing", they made a corpus of audio data talk taken from a serial film after Tepperman used a telephone dialog corpus taken from Switchboard and Fisher [3]. Video data must go through a sarcasm screening process to take part of the voice that contains sarcasm and determine whether the piece that was taken is the part of the voice that was taken only as sarcasm or a piece of conversation to take context. Video data is then converted to voice version only.

Data Text. Social media statistics from 2019 show that there are 3.2 billion social media users worldwide, and this number is only growing. That equates to about 42% of the current population. Every second, on average, around 6,000 tweets are tweeted on Twitter [1]. Besides, Twitter also provides an API that allows you to do complex queries like

[1] https://www.internetlivestats.com/twitter-statistics/

pulling every tweet about a certain topic within the last twenty minutes or pull a certain user's non-retweeted tweets[2]. This opens great opportunities for data miners, sentiment analyzers, digital marketers, and all fields that need data related to sentiment analysis, etc. It can be seen from Table 1 that the number of sarcasm detection studies using Twitter data is 60% more than text data from other sites. Indeed, Twitter is a paradise for data, but it is not just Twitter. There is also much data that is widespread on the world wide web, such as quote snippet data taken from sites conducted by source: Are word embedding.

Text datasets can be divided into three types: short text, long text, and other. Short text is a text dataset that only contains utterances that are considered sarcasm so that the number of words in a sarcasm sentence tends to be short. This is often found in data tweets that have a limit of 140 words per tweet. Simply put, long text can be described as text data that has several sentences longer than short text, such as conversation, discussion, and review. The benefit of long text is flexibility to understand its context. In tweets, context can be captured through hashtags. The application of sarcasm detection through a hashtag is considered as short text in [4]. However, referring to the main understanding of short text, hashtag cannot be considered fully as short text because hashtag is not a sentence, but a word that contains or has the possibility as sarcasm.

Table 1. Sarcasm Detection in Different Data Types

Year	Type	Data	Accuracy	Resource
2006	Audio	Telephone Dialogues from Switchboard and Fisher corpora	87% contextual + spectral	[3]
2010	Text	Twitter	90% all no enrich	[5]
2013	Audio	Daria, Animated Television Show from MTV	81.57% baseline + unigram + intenbigrams	[6]
2015	Text	Twitter	88.9%	[7]
2015	Text	Twitter	98% PP+POS with Bayesian Model Averaging	[8]
2016	Text	Quotes from Books gathered through website	97.8% Dep. Wt.	[9]
2016	Text	Twitter	77.11% on 6th test with SVM	[10]
2016	Text	Twitter	83.1% with Rand. Forest	[11]
2019	Multimodal	Serial Film from Youtube	72% precision, speaker dependent with best text and video features	[12]
2019	Text	Twitter	91%	[13]

2 https://chatbotslife.com/twitter-data-mining-a-guide-to-big-data-analytics-using-python-4efc8ccfa219

B. Approaches

Most studies use the same steps in detecting sarcasm, which are all structured from choosing dataset, pre-processing, feature extraction, and classifying sarcasm. In audio-based sarcasm detection, the first step is to determine how data will be processed so that data collection can be conducted well. There are not many techniques in audio-based data preprocessing since there are only a few studies that discuss sarcasm detection using audio-based data. In text-based sarcasm detection, for example, sarcasm detection is performed through text mining where the first step is to gather the data as mentioned at text dataset types. Several techniques are used to prepare clean text data in preprocessing step, such as tokenization to split sentences into tokens, removal of meaningless frequent stop words, conflating forms of words with stemming [14]. All of the key features from the data will be extracted before it classifies whether it is sarcasm or not.

Most works rely on Support Vector Machines (SVM), the most popular statistical method [13, 32, 38, 56, 67, 68]. Several works combine it with another method as an effort to maximize accuracy. The study in [15] got SVM-HMM 84.2 F-Score higher than SEARN in Sequence Labelling better than classification algorithm for conversational data. Recently, as the Deep Learning method gains more popularity along with the robust development of Big Data, the technique gets attention in the sarcasm detection area. A few works have been reported using deep learning for automatic sarcasm detection [9], [16].

© Copyright: All rights reserved.

The performance achieved in [17] improves by combining Convolutional Neural Network, Deep Neural Network and Recurrent Neural Network with LSTM where it can plot long term dependencies by defining each memory cell with a set of gates < d. Combining these three Neural Network techniques has a better performance compared to recursive SVM that has been modeled like in Fig 1.

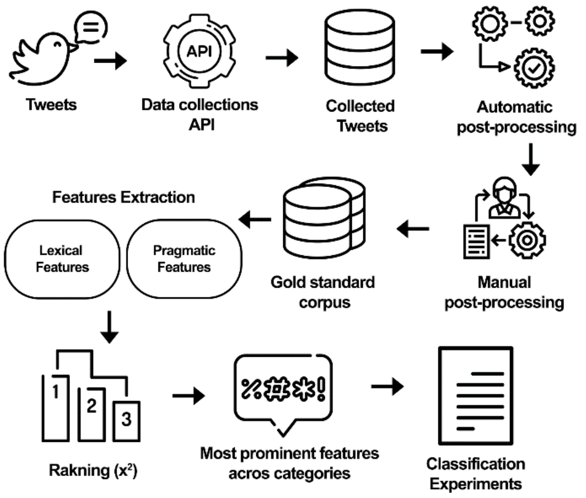

Fig 1. Computational Framework for Sarcasm Detection [10]

C. Parameters

From the outline, each work has different method. Mostly, it is caused by the variety of parameters used. This variety includes a big portion of parameters, which can be the context, features, or even the approach constraints that can be used. These parameters have a strong correlation between each other. It determines the architecture of the method.

Approaches Rules. The name approaches rules act as determinant in the classifying process. Some works use hashtag as the main key indicator of sarcasm. They developed an algorithm to split tokens from the hashtag with a language modeling approach based on unigram/bigram frequencies for the ambiguity problem [18]. The problem of hashtag-based may lead to quality degradation of the training data for reasons such as incorrect use of a sarcasm-indicative hashtag like the hashtag '#sarcastic' can be identified as sarcastic but in fact, it is not a sarcasm tweet. Besides hashtag, there is a rule-based approach to classify sarcasm where a sentence has a positive verb and negative situation phrase. They made seven POS bigram patterns: V+V, V+ADV, ADV+V, "to"+V, V+NOUN, V+PRO, V+ADJ with a bootstrapped model lists which can recognize sarcasm where the SVM classifier misses [19].

Context. The term "context" here refers to any information beyond the text to be predicted. This context is mostly found in text-based data which can be divided into three types: authors-specific context, topical context, and conversational context. The presence of these contexts is expected to improve automatic sarcasm detection with additional data. Some context is dependent on who the author is, in this case, the information of the author has a benefit to retrieve extra information about the author. In the example, the expression "I love Math!" could be ambiguous. With the historical tweets of the author, it could be used as a feature to classify sarcasm. Various historical tweets of the author, like the hashtag, language and detected sarcasm, could be considered as author-specific context too [20]. Profile information, historical sentiment is counted as an author-specific context too. Another type of context is topical context, which is a context with a principle that some topics are more likely to contain sarcasm than other topics. The study in [21] presents a sarcasm topic model that uses sentiment mixture in tweets to discover sarcasm-prevalent topics. They note how topics like elections are more likely to evoke sarcasm as compared to funeral or fathers' day. Conversational context requires long text data, which is a conversational data. Conversational context helps when a sentence does not have any negative situation. The study in [22] reveals the concatenation of the previous post in a discussion forum thread along with the target post leads to an improvement in precision.

III. RESULTS AND DISCUSSIONS

Based on all the latest methods and approaches about automatic sarcasm detection, there is not much development in the speech area. People seem more interested in processing the textual data because the amount and the variety of textual data make it easy to gather and explore the method of automatic sarcasm detection. In processing textual data, the hashtag-based rule may be used with another context to improve the accuracy of automatic sarcasm detection, but hashtag-based rule alone is not reliable due to the bias of specifically used hashtag that does not represent sarcasm like "#sarcasm" and "#politic". On the other hand, the hashtag-based rule might be useful to categorize related topic to support the use of topical context and author-specific context. Topical context holds a new key in improving the accuracy of automatic sarcasm detection, while the author-specific still has a disadvantage when the tweets come from new and private accounts.

Table 2. Experimental Results [17]

Model	Feature/Hyper Parameter	Precision	Recall	F-score
LSTM + LSTM	Hidden memory unit = 64	.849	.816	.832
	Hidden memory unit = 128	.854	.871	.862
	Hidden memory unit = 256	.868	.89	.879
CNN + LSTM + DNN (with dropout)	filter size = 256 + filter width = 2 + HMU = 256	.899	.91	.904
CNN + LSTM + DNN	filter size = 256 + filter width = 2 + HMU = 256	.912	.911	.912
CNN + LSTM + DNN	filter size = 256 + filter width = 3 + HMU = 256	.919	.923	.921

Retrieval Number: D7215118419/2019©BEIESP
DOI:10.35940/ijrte.D7215.118419
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)
© Copyright: All rights reserved.

From this research, we found that the latest research using *Recurrent Neural Network* (RNN) works better in extraction features in revealing context which contributes significant number to the accuracy, which reaches its best to 91% with 3 RNN layers and 256 *Long Term Short Memor*y (LSTM) cells as referred to Table 2. Using more hidden memory unit leads to better accuracy.

IV. CONCLUSIONS

Considering the discovery, the variety of datasets, approaches and parameter, each of them has an advantage and disadvantage. Use of single context could lead to a biased result. Every possible context should be considered to be combined to reach a maximum result. Since sarcasm is very context-based and depends on situation and time, there is a lot of textual contexts that can be used in acoustic context when the textual data itself cannot retrieve any context information from acoustic data, such as topical context. The work that has been conducted in the area of acoustic data is still limited. We found that sarcasm detection through acoustic data has the potential to reach better accuracy since it can be extracted into two data types. The acoustic feature which reveals the tonal information of a speech and textual feature which can be extracted with speech-to-text technology. From this research, we also realize that the common textual method cannot be applied since the extracted textual data from speech does not have any hashtag or author historical tweets. We also found that in future work, the dependencies of these contexts cannot be applied in a universal data and tools since it needs non-universal features like hashtag and user historical tweets. It means if we want to apply this concept to a real-time automatic sarcasm detection, let's say in a mobile phone, we do not use any hashtag while we were saying a sarcasm.

REFERENCES

1. S. K. Bharti, B. Vachha, R. K. Pradhan, K. S. Babu, and S. K. Jena, "Sarcastic sentiment detection in tweets streamed in real time: a big data approach," Digit. Commun. Networks, vol. 2, no. 3, pp. 108–121, 2016.
2. A. Mishra, D. Kanojia, S. Nagar, K. Dey, and P. Bhattacharyya, "Harnessing cognitive features for sarcasm detection," 54th Annu. Meet. Assoc. Comput. Linguist. ACL 2016 - Long Pap., vol. 2, pp. 1095–1104, 2016.
3. J. Tepperman, D. Traum, and S. Narayanan, "Yeah right : Sarcasm recognition for spoken dialogue systems," no. January, 2006.
4. A. Joshi, P. Bhattacharyya, and M. J. Carman, "Automatic sarcasm detection: A survey," ACM Comput. Surv., vol. 50, no. 5, 2017.
5. D. Davidov, O. Tsur, and A. Rappoport, "Semi-supervised recognition of sarcastic sentences in twitter and Amazon," CoNLL 2010 - Fourteenth Conf. Comput. Nat. Lang. Learn. Proc. Conf., no. July, pp. 107–116, 2010.
6. R. Rakov and A. Rosenberg, "' Sure , I Did The Right Thing ': A System for S arcasm Detection in Speech Linguistics Department , The Graduate Center CUNY , New York , USA Computer Science Department , Queens College CUNY , New York , USA," Interspeech, no. August, pp. 842–846, 2013.
7. D. Ghosh, W. Guo, and S. Muresan, "Sarcastic or not: Word embeddings to predict the literal or sarcastic meaning of words," Conf. Proc. - EMNLP 2015 Conf. Empir. Methods Nat. Lang. Process., no. September, pp. 1003–1012, 2015.
8. E. Fersini, F. A. Pozzi, and E. Messina, "Detecting irony and sarcasm in microblogs: The role of expressive signals and ensemble classifiers," Proc. 2015 IEEE Int. Conf. Data Sci. Adv. Anal. DSAA 2015, 2015.
9. A. Joshi, V. Tripathi, K. Patel, P. Bhattacharyya, and M. Carman, "Are Word Embedding-based Features Useful for Sarcasm Detection?," no. 2013, pp. 1006–1011, 2016.
10. S. Muresan, R. Gonzalez-Ibanez, D. Ghosh, and N. Wacholder, "Identification of nonliteral language in social media: A case study on sarcasm," J. Assoc. Inf. Sci. Technol., vol. 67, no. 11, pp. 2725–2737, 2016.
11. M. Bouazizi and T. Otsuki, "A Pattern-Based Approach for Sarcasm Detection on Twitter," IEEE Access, vol. 4, no. November, pp. 5477–5488, 2016.
12. S. Castro, D. Hazarika, V. Pérez-Rosas, R. Zimmermann, R. Mihalcea, and S. Poria, "Towards Multimodal Sarcasm Detection (An _Obviously_ Perfect Paper)," 2019.
13. S. Porwal, G. Ostwal, A. Phadtare, M. Pandey, and M. V. Marathe, "Sarcasm Detection Using Recurrent Neural Network," Proc. 2nd Int. Conf. Intell. Comput. Control Syst. ICICCS 2018, no. Iciccs, pp. 746–748, 2019.
14. C. Paper, "Preprocessing Techniques for Text Mining Preprocessing Techniques for Text Mining," no. October 2014, 2016.
15. A. Joshi, V. Tripathi, P. Bhattacharyya, and M. J. Carman, "Harnessing Sequence Labeling for Sarcasm Detection in Dialogue from TV Series `Friends'," no. August, pp. 146–155, 2016.
16. S. Amir, B. C. Wallace, H. Lyu, P. Carvalho, and M. J. Silva, "Modelling Context with User Embeddings for Sarcasm Detection in Social Media," pp. 167–177, 2016.
17. A. Ghosh and D. T. Veale, "Fracking Sarcasm using Neural Network," pp. 161–169, 2016.
18. D. Maynard and M. A. Greenwood, "Who cares about sarcastic tweets? Investigating the impact of sarcasm on sentiment analysis," Proc. 9th Int. Conf. Lang. Resour. Eval. Lr. 2014, pp. 4238–4243, 2014.
19. E. Riloff, A. Qadir, P. Surve, L. De Silva, N. Gilbert, and R. Huang, "Sarcasm as Contrast between a Positive Sentiment and Negative Situation," EMNLP 2013 - 2013 Conf. Empir. Methods Nat. Lang. Process. Proc. Conf., no. October, pp. 704–714, 2013.
20. A. Rajadesingan, R. Zafarani, and H. Liu, "Sarcasm Detection on Twitter," pp. 97–106, 2015.
21. A. Joshi, P. Jain, P. Bhattacharyya, and M. Carman, "`Who would have thought of that!': A Hierarchical Topic Model for Extraction of Sarcasm-prevalent Topics and Sarcasm Detection," 2016.
22. A. Joshi, V. Sharma, and P. Bhattacharyya, "Harnessing context incongruity for sarcasm detection," ACL-IJCNLP 2015 - 53rd Annu. Meet. Assoc. Comput. Linguist. 7th Int. Jt. Conf. Nat. Lang. Process. Asian Fed. Nat. Lang. Process. Proc. Conf., vol. 2, no. 2003, pp. 757–762, 2015.

AUTHORS PROFILE

Steve Michael Ruru A student of Computer Science in Bina Nusantara University (BINUS) in an accelerated Master Degree program. His first interest in the field of speech technology and the connection with psychology started in 2018. He is currently working in the web development concerned with UX field. He lives as a citizen of Indonesia in Jakarta.

Amalia Zahra obtained her PhD degree from the School of Computer Science and Informatics, University College Dublin (UCD), Ireland in 2014, and her Bachelor degree from the Faculty of Computer Science, University of Indonesia in 2008. She does not have a Master degree. After finishing her Bachelor degree in 2008, she worked as a research assistant for a year, and then continued to pursue PhD immediately afterwards. The topic of her PhD thesis was about spoken language identification. Her research interest has been in the area of speech technology, machine learning, and computational linguistics since 2008. Therefore, her publications are related to such topics. Currently, she is working as a lecturer at the Master of Information Technology, Binus University, Jakarta, Indonesia.

Retrieval Number: D7215118419/2019©BEIESP
DOI:10.35940/ijrte.D7215.118419
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)
© Copyright: All rights reserved.