

Sarcasm Detection and Building an English Language Corpus in Real Time

Oliver Cakebread-Andrews
University of Wolverhampton
o.p.cakebread-andrews@wlv.ac.uk

Abstract

This is a research proposal for doctoral research into sarcasm detection, and the real-time compilation of an English language corpus of sarcastic utterances. It details the previous research into similar topics, the potential research directions and the research aims.

1 Introduction

Within the discipline of computational linguistics, and specifically in the field of sentiment analysis, Sarcasm Detection is a fast-growing area of research. In the past twelve months alone there have been multiple papers written on the subject of sarcasm and irony detection (Aboobaker & Illavarasan, 2020; Avvaru et al., 2020; Buzea et al., 2020; Eke et al., 2021). Sarcasm Detection uses Natural Language Processing models along with sentiment analysis based on linguistic cues such as hyperbole or polarity shifts, as well as context to decode whether or not an utterance is sarcastic. Pelsler & Murrell (2019) give the example of “*Great! I love waking up sick!*”, which is easy for a human to detect as sarcastic due to “strong polarity shifts”, but would remain a challenge for a machine. Suggested applications for research into sarcasm detection include improving the understanding of users’ or customers’ emotions and opinions about policies, products or services for businesses and governments or policy-makers (Eke et al., 2021; Vyas & Uma, 2019).

2 Background and Literature Review

2.1 Data Collection

In order to obtain data of sarcastic utterances for corpora used in algorithmic training, many studies used social media (Baruah et al., 2020; Bharti, 2019) such as Reddit (Ghosh et al., 2020; Mishra, Kaushik, & Dey, 2020) and Twitter (Avvaru et al.,

2020; Ghosh et al., 2020). This provides a wealth of naturally generated, authentic data that is often already tagged as sarcastic (hashtags in the case of Twitter and /s in the case of Reddit), making it easy to collect and categorise. This study would similarly utilise Reddit to develop a novel corpus of sarcastic utterances for the purpose of training an algorithm to detect sarcasm. It appears that in the vast majority of cases, researchers use either an existing corpus of sarcastic utterances (Ghosh et al., 2018; Ghosh et al., 2020) such as the one compiled by Khodak et al. (2018), or more commonly according to Eke et al. (2019), they gather their own unique data to use just in their study (Avvaru et al., 2020; Ghosh et al., 2020; Ren et al., 2018). These corpora generally tend to be considerably smaller (e.g. 4000 Tweets (Ghosh et al., 2020), 5000 Tweets (Jaiswal, 2020), 39000 Tweets (Ghosh & Veale, 2016), 900 Tweets (González-Ibáñez et al., 2011)) though this could be due in part to Twitter’s restrictions on how many Tweets can be freely downloaded through their API. Although there are exceptions such as Kunneman et al. (2015) who created a corpus of 406 thousand Dutch Tweets. Indeed, it has been noted as recently as 2021 (Eke et al., 2021) that there is a lack of training data for classifiers. It seems that there have not been any major attempts to create a dynamic corpus that continues to automatically detect sarcastic utterances and update the data by adding to the corpus, and consequently learning from the additions. Therefore, the first aim of this study would be to compile a corpus of sarcastic data from Reddit, as well as potentially other sources, that can be used to train an algorithm to detect additional sarcastic utterances in real-time and continuously update, and learn from, the corpus.

2.2 Methodology

Various methodologies and learning models have been employed in to detect sarcasm and irony including, but not limited to, the more

traditional models such as Support Vector Machines (SVM) (Castro, et al., 2019; González-Ibáñez et al., 2011) and more modern neural network models like Long Short Term Memory (LSTM) and Bi-Directional LSTM (Avvaru et al., 2020; Ghosh et al., 2018; Ghosh et al., 2020), Convolutional Neural Networks (CNN), (Castro, et al., 2019; Jaiswal, 2020), Deep Neural Networks (DNN) (Ghosh & Veale, 2016) and Hierarchical Fusion Model (Pelser & Murrell, 2019). CNN and LSTM have been by far the most popular methods for approaching sarcasm detection in the past decade. However, many of the more recent studies have tended to use Bidirectional Encoder Representations from Transformers (BERT) (Avvaru et al., 2020; Baruah et al., 2020) or suggested using BERT in future work (Eke et al., 2021). Therefore, this study is likely to use LSTM, Bi-LSTM, CNN or BERT models for analysis, or a combination of multiple models as has been seen in other papers (Kumar & Garg, 2019; Zhang et al., 2019)

Current accuracy rates of sarcasm detection models can vary quite widely depending on a multitude of factors such as models used, training data and what is being tested. Focussing on the more recent studies, which can be assumed to take advantage of the most modern methods for increased accuracy, there are studies using Bi-LSTM with F-scores from the high 70s (Kumar & Garg, 2019; Mishra et al., 2020) for Reddit, to the low 90s (Mishra et al., 2020) for Twitter, low to high 70s for studies using BERT models (Avvaru et al., 2020; Jaiswal, 2020) and mid 80s for multimodal studies using LSTM (Cai et al., 2019) and slightly lower for BERT (Xu et al., 2020). In the 2020 Sarcasm Detection Shared Task (2020), Ghosh, Vajpayee and Muresan reported findings ranging from 59 to 83 on Reddit and 68 to 93 on Twitter. Studies on Twitter seem to consistently outperform Reddit. Whether this is because comments on Reddit can be much longer, there has been more research conducted on Twitter (as evidenced in the second paragraph) or whether there is some other reason is unclear. Due to the greater opportunity for improvement, the initial focus of this study will be on Reddit, but there is certainly a possibility of expanding to Twitter as well.

Whilst all the studies are researching sarcasm or irony detection, the main focus of analysis varies according to the study, such as semantics (Xu et al., 2020), hyperbole (Kunneman et al., 2015) emojis (Wiguna et al., 2021), and taking the utterance with (Jaiswal, 2020; Baruah et al., 2020)

or without context. There have also been several studies done into multimodal sarcasm detection (Castro, et al., 2019; Pelser & Murrell, 2019; Cai et al., 2019). Limitations with current multimodal studies seem to be small corpora (Castro, et al., 2019), or the same corpora (Xu et al., 2020), lack of consideration of multiple participants (Wang et al., 2020) and lack of audio (Cai et al., 2019; Castro, et al., 2019; Pelser & Murrell, 2019). Due to social media posts often taking advantage of multiple formats, extending the scope of this future study to include multimodal analysis will be considered at a later stage in the research.

2.3 Languages

The overwhelming majority of studies continue to be based on English, but other languages are emerging as subjects of sarcasm detection, such as Dutch (Kunneman et al., 2015), Spanish (Frenda & Patti, 2019; Ortega-Bueno et al., 2019), Romanian (Buzea et al., 2020), Arabic (Ranasinghe et al., 2019), Hindi (Jain et al., 2020). Despite this, there are still many languages with negligible or no studies of sarcasm detection done in that medium, leaving the opportunity to transfer the findings from the English study onto these other languages. Japanese for example seems to have very few studies published related to sarcasm detection (Hiai & Shimada, 2018). While there are plenty of examples of studies into cross-linguistics transfer learning, there appears to be very few that focus on, or even mention, sarcasm (Chronopoulou et al., 2019; Tavan et al., 2020). Moving further away from Germanic and Romance languages and cultures may mean limited results, however it is a further aim of this study to discover what patterns and similarities can be found when transferring results to a second language.

3 Aims of the Study

To summarise the aims and focus of this research, there are two current aims which will be the main focus of the research. They are as follows:

1. To compile a large corpus of English sarcastic utterances that will be updated in real-time using social media
2. To use a combination of natural language processing techniques and deep neural networks to identify sarcasm with high accuracy

The first aim is to try and tackle the issue of a lack of large and up-to-date corpora to provide training data for detection algorithms and models. The second is to attempt to improve the current detection rates, particularly on natural data that is unrelated to the training set.

There are also two potential further research directions that may be undertaken depending on the eventual feasibility of introducing other areas of research. These aims are the following:

1. To expand the scope of the study to multimodal analyses to take into consideration images and potentially audio
2. To use transfer learning to transfer findings to a non-European language such as Japanese

These additional directions would be to attempt to expand into areas of sarcasm detection that currently have had little to no research conducted into them.

References

- Aboobaker, J., & Illavarasan, E. (2020). A Survey on Sarcasm Detection and Challenges. *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*. <https://doi.org/10.1109/ICACCS48705.2020.9074163>
- Avvaru, A., Vobilisetty, S., & Mamidi, R. (2020). Detection Sarcasm in Conversation Context Using Transformer-Based Models. *Proceedings of the Second Workshop on Figurative Language Processing*. <http://dx.doi.org/10.18653/v1/2020.figlang-1.15>
- Baruah, A., Das, K., Barbhuiya, F., & Dey, K. (2020). Context-Aware Sarcasm Detection Using BERT. *Proceedings of the Second Workshop on Figurative Language Processing*. <http://dx.doi.org/10.18653/v1/2020.figlang-1.12>
- Bharti, S. K. (2019). *Sarcasm Detection in Textual Data: A Supervised Approach*. Doctoral Dissertation. <https://doi.org/10.4018/978-1-5225-7186-5.CH010>
- Buzea, M., Trăușan-Matu, S., & Rebedea, T. (2020). Automatic Irony Detection for Romanian Online News. *24th International Conference on System Theory, Control and Computing*. <https://doi.org/10.1109/ICSTCC50638.2020.9259715>
- Cai, Y., Cai, H., & Wan, X. (2019). Multi-Modal Sarcasm Detection in Twitter with Hierarchical Fusion Model. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. <http://dx.doi.org/10.18653/v1/P19-1239>
- Castro, S., Hazarika, D., Pérez-Rosas, V., Zimmermann, R., Mihalcea, R., & Poria, S. (2019). Towards Multimodal Sarcasm Detection (An_Obviously_Perfect Paper). *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. <http://dx.doi.org/10.18653/v1/P19-1455>
- Chronopoulou, A., Baziotis, C., & Potamianos, A. (2019). An Embarrassingly Simple Approach for Transfer Learning from Pretrained Language Models. *Proceedings of the 2019 Conference of the North*. <http://dx.doi.org/10.18653/v1/N19-1213>
- Eke, C. I., Norman, A. A., & Shuib, L. (2021). Multi-feature fusion framework for sarcasm identification on twitter data: A machine learning based approach. *PLOS ONE*, 16(6). <https://doi.org/10.1371/journal.pone.0252918>
- Eke, C. I., Norman, A. A., Shuib, L., & Nweke, H. F. (2019). Sarcasm identification in textual data: systematic review, research challenges and open directions. *Artificial Intelligence Review*, 53(6), 4215-4258. <https://doi.org/10.1007/s10462-019-09791-8>
- Frenda, S., & Patti, V. (2019). Computational models for irony detection in three spanish variants. *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2019), co-located with 34th Conference of the Spanish Society for Natural Language Processing (SEPLN 2019)*. CEUR. <http://hdl.handle.net/2318/1795396>
- Ghosh, A., & Veale, D. T. (2016). Fracking Sarcasm using Neural Network. *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media*

- Analysis*.
<http://dx.doi.org/10.18653/v1/W16-0425>
- Ghosh, D., Fabbri, A. R., & Muresan, S. (2018). Sarcasm Analysis Using Conversation Context. *Computational Linguistics*, 44(4), 755-792.
https://doi.org/10.1162/coli_a_00336
- Ghosh, D., Vajpayee, A., & Muresan, S. (2020). A Report on the 2020 Sarcasm Detection Shared Task. *Proceedings of the Second Workshop on Figurative Language Processing*.
<http://dx.doi.org/10.18653/v1/2020.figlang-1.1>
- González-Ibáñez, R., Muresan, S., & Wacholder, N. (2011). Identifying sarcasm in twitter: a closer look. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, (pp. 581-596).
<https://aclanthology.org/P11-2102>
- Hiai, S., & Shimada, K. (2018). Sarcasm Detection Using Features Based on Indicator and Roles. *Advances in Intelligent Systems and Computing*, 418-428.
https://doi.org/10.1007/978-3-319-72550-5_40
- Jain, D., Kumar, A., & Garg, G. (2020). Sarcasm detection in mash-up language using soft-attention based bi-directional LSTM and feature-rich CNN. *Applied Soft Computing*, 91.
<https://doi.org/10.1016/j.asoc.2020.106198>
- Jaiswal, N. (2020). Neural Sarcasm Detection using Conversation Context. *Proceedings of the Second Workshop on Figurative Language Processing*.
<http://dx.doi.org/10.18653/v1/2020.figlang-1.11>
- Khodak, M., Saunshi, N., & Vodrahalli, K. (2018). A Large Self-Annotated Corpus for Sarcasm. *ArXiv*. <https://aclanthology.org/L18-1102>
- Kumar, A., & Garg, G. (2019). Sarcasm Detection Using Feature-Variant Learning Models. *Proceedings of ICETIT 2019*, (pp. 683-693).
https://doi.org/10.1007/978-3-030-30577-2_61
- Kunneman, F., Liebrecht, C., Mulken, M. V., & Bosch, A. V. (2015). Signalling sarcasm: From hyperbole to hashtag. *Information Processing & Management*, 51(4), 500-509.
<https://doi.org/10.1016/j.ipm.2014.07.006>
- Mishra, P., Kaushik, S., & Dey, K. (2020). Bi-ISCA: Bidirectional Inter-Sentence Contextual Attention Mechanism for Detecting Sarcasm in User Generated Noisy Short Text. *arXiv*.
<https://arxiv.org/abs/2011.11465>
- Ortega-Bueno, R., Rangel, F., Hernández Fariás, D., Rosso, P., Montes-y-Gómez, M., & Medina Pagola, J. E. (2019). Overview of the task on irony detection in Spanish variants. *Proceedings of the Iberian languages evaluation forum (IberLEF 2019), co-located with 34th conference of the Spanish Society for natural language processing (SEPLN 2019)*. CEUR.
- Pelser, D., & Murrell, H. (2019). Deep and Dense Sarcasm Detection. *ArXiv*.
<https://arxiv.org/abs/1911.07474>
- Ranasinghe, T., Saadany, H., Plum, A., Mandhari, S., Mohamed, E. O., & Mitkov, R. (2019). RGCL at IDAT: Deep Learning models for Irony Detection in Arabic Language. *Proceedings of the IDAT@FIRE2019*. CEUR Workshop Proceedings.
<http://hdl.handle.net/2436/622828>
- Ren, Y., Ji, D., & Ren, H. (2018). Context-augmented convolutional neural networks for twitter sarcasm detection. *Neurocomputing*, 308, 1-7.
<https://doi.org/10.1016/j.neucom.2018.03.047>
- Tavan, E., Rahmati, A., & Ali Keyvanrad, M. (2020). Persian Emoji Prediction Using Deep Learning and Emoji Embedding. *2020 10th International Conference on Computer and Knowledge Engineering (ICCKE)*.
<https://doi.org/10.1109/ICCKE50421.2020.9303639>
- Vyas, V., & Uma, V. (2019). Approaches to Sentiment Analysis on Product Reviews. *Advances in Business Information Systems and Analytics*, 15-30.
<https://doi.org/10.4018/978-1-5225-4999-4.CH002>
- Wang, X., Sun, X., Yang, T., & Wang, H. (2020). Building a Bridge: A Method for Image-Text Sarcasm Detection Without Pretraining on Image-Text Data. *Proceedings of the First International Workshop on Natural*

Language Processing Beyond Text.
<http://dx.doi.org/10.18653/v1/2020.nlpbt-1.3>

- Wiguna, B. S., Hudiyanti, C. V., Alqis Rausanfitra, A., & Arifin, A. Z. (2021). Sarcasm Detection Engine for Twitter Sentiment Analysis using Textual and Emoji Feature. *Jurnal Ilmu Komputer Dan Informasi*, 41(1), 1-8.
<https://doi.org/10.21609/jiki.v14i1.812>
- Xu, N., Zeng, Z., & Mao, W. (2020). Reasoning with Multimodal Sarcastic Tweets via Modeling Cross-Modality Contrast and Semantic Association. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics.*
<http://dx.doi.org/10.18653/v1/2020.acl-main.349>
- Zhang, S., Zhang, X., Chan, J., & Rosso, P. (2019). Irony detection via sentiment-based transfer learning. *Information Processing & Management*, 56(5), 1633-1644.
<https://doi.org/10.1016/j.ipm.2019.04.006>