Sarcasm Detection

Diana Abagyan
dabagyan@uw.edu

Libby Merchant
emercha@uw.edu

Emma McKibbin
echm@uw.edu

Jade Phoreman
jphore@uw.edu

Catherine Ball
catball@uw.edu

Abstract

1 Introduction

2 Task Description

Our primary task is sarcasm detection using a model trained on the SARC dataset (Khodak et al., 2018) found here and characterized by the following dimensions:

- **Affect type**: emotion (sarcasm)
- **Recognition type**: classification
- **Genre**: Reddit posts
- **Target**: N/A
- **Modality**: text
- **Language**: English

The SARC dataset was collected from Reddit posts and contains 1.3 million sarcastic statements and many more non-sarcastic statements each with author, topic, and context. Because the data instances are obtained from Reddit posts, the data is self-annotated, meaning that the author labeled sarcastic comments using the tone indicator /s. There is some noise associated with self-annotation using the sarcasm tone indicator which is discussed in Khodak et al. (2018). The SARC dataset is also unique among other sarcasm datasets because it is unbalanced, meaning there are significantly more instances of non-sarcasm than sarcasm.

Our adaptation task is sarcasm detection using a model trained on a dataset we will collect, from a significantly higher proportion of neurodivergent users. We will collect text posts from subreddits such as r/Neurodivergent, r/neurodiversity, and r/autism, many of which will be self-labeled as sarcastic with the /s tone indicator. There does not seem to be any literature on using computational models to perform sarcasm detection among neurodivergent populations specifically, so instead we refer to adjacent literature for guidance, such as Au-Yeung et al. (2015) and Febiana Christanti et al. (2022).

Previous studies do not agree as to whether ASD (Autism Spectrum Disorder) individuals perform worse than neurotypical individuals when processing irony. In an effort to disambiguate the fields' past findings, Au-Yeung et al. (2015) conducted an eye-tracking study on ASD and TD (typically-developing) individuals. In their experiment, they showed participants short passages which contained a sentence that could be either ironic or non-ironic, depending on the surrounding context. For some passages, a comprehension question was presented to test whether the participant had correctly interpreted the presence or absence of irony. They found both groups performed similarly in terms of accuracy and detection time; however, the ASD group took more time to read all passages, regardless of irony. The authors suggest that, regardless of whether there is a processing deficiency or not, this finding shows that ASD individuals have lower confidence in their ability to detect irony.

Text on social media does not have paralinguistic cues such as tone of voice, facial expressions, and body language which would help disambiguate sarcasm. Tone indicators have emerged in neurodiverse communities online in order to compensate (Febiana Christanti et al., 2022).

Our evaluation methodology for both tasks will involve comparing the correct class labels of "sarcasm" and "non-sarcasm" to the predicted class labels given by the model. We will use metrics such as f1-score, recall, and precision, as Khodak et al. did when training and testing sarcasm detection models using the SARC dataset.

3 System Overview

4 Approach

5 Results

6 Discussion

7 Ethical Considerations

8 Conclusion

References

Sheena K. Au-Yeung, Johanna K. Kaakinen, Simon P. Liversedge, and Valerie Benson. 2015. Processing of written irony in autism spectrum disorder: An eye-movement study. *Autism Research*, 8(6):749–760.

Maria Febiana Christanti, Puri Bestari Mardani, and Khansa Ayu Fadhila. 2022. Analysing the meaning of tone indicators by neurodivergent community in twitter. *International Journal of Social Science Research and Review*, 5(1):5–15.

Mikhail Khodak, Nikunj Saunshi, and Kiran Vodrahalli. 2018. A large self-annotated corpus for sarcasm. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).