

Sarcasm Classification

Amy Tzu-Yu Chen
University of Washington
amy17519@uw.edu

David Roesler
University of Washington
droesl@uw.edu

Diana Baumgartner-Zhang
University of Washington
diazhang@uw.edu

Juliana McCausland
University of Washington
jumc1469@uw.edu

Abstract

TBD.

1 Introduction

[DRAFT] In any text format, detecting whether someone is sarcastic or serious can be challenging. For spoken interactions, we can rely on intonation, facial expressions, and other non-verbal cues to inform us on the underlying meanings. On Twitter or other social media platforms, users often express their opinions or thoughts solely through text. Even for other people, discerning whether a text should be interpreted literally or figuratively can prove to be a challenge.

Sarcasm differs from irony in that it is mocking and often critical in nature. Often, the underlying sentiment is negative, but appears positive on the surface.

The use of sarcasm on the internet is ubiquitous, and its presence can disrupt computational systems of sentiment analysis [Liu, 2010](#).

2 Task Description

Our primary task is sarcasm detection in English Twitter text, which we treat as a binary classification problem. Given a text, our task is to determine whether it is sarcastic or non-sarcastic. Sarcasm is a form of verbal irony through which a speaker expresses their stance toward a topic, which often takes the form of contempt or derogation [Wilson, 2006](#). Automatic sarcasm detection [Joshi et al., 2017](#) is the prediction of the presence of sarcasm in text. Twitter, a platform often used to express the critical viewpoints of its users, has been a common data source for sarcasm detection models [Sarsam et al., 2020](#). To train and evaluate our model, we make use of the Twitter sarcasm dataset from SemEval 2022 Task 6, iSarcasmEval [Oprea and Magdy, 2020](#). Unlike sarcasm datasets labeled by third-party annotators, this dataset contains labels

provided by the authors of the tweets themselves. The iSarcasmEval data includes both English and Arabic sets.

As our secondary adaptation task, we perform sentiment detection on the Arabic portion of the dataset. To evaluate the performance of our binary classification model, we measure F1 score on the positive (sarcastic) class.

3 System Overview

TBD.

4 Approach

TBD.

5 Results

TBD.

6 Conclusion

TBD.

References

- Aditya Joshi, Pushpak Bhattacharyya, and Mark J. Carman. 2017. [Automatic sarcasm detection: A survey](#). *ACM Comput. Surv.*, 50(5).
- Bing Liu. 2010. Sentiment analysis and subjectivity. In *Handbook of Natural Language Processing*.
- Silviu Oprea and Walid Magdy. 2020. [iSarcasm: A dataset of intended sarcasm](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1279–1289, Online. Association for Computational Linguistics.
- Samer Muthana Sarsam, Hosam Al-Samarraie, Ahmed Ibrahim Alzahrani, and Bianca Wright. 2020. [Sarcasm detection using machine learning algorithms in twitter: A systematic review](#). *International Journal of Market Research*, 62(5):578–598.
- Deirdre Wilson. 2006. [The pragmatics of verbal irony: Echo or pretence?](#) *Lingua*, 116(10):1722–1743. Language in Mind: A Tribute to Neil Smith on the Occasion of his Retirement.