Detecting Sarcasm in Text: Unraveling the Mystery hidden in text

Michael Kalish , Mahmoud Ghanem , Eshan Bhatnagar

Objective:

The objective is to use a range of small BERT models in combination with a diverse set of embeddings (custom and default) to explore whether it is possible to improve on the text classification of sarcastic tweets with less compute and higher degrees of customization.

Importance & Challenges

Sarcasm is obvious 🤪

A model that can effectively detect sarcasm has multiple, important applications:

- Help neurodivergent people identify sarcastic notes and cues to better navigate online social environments
- Assist people who may be wrongfully banned or flagged for threatening or offensive speech on various platforms who may have only been joking/sarcastic and not actually intended any harm
- Customer Feedback interpretation
- Detecting sarcasm in online text especially in social media content, where sarcasm is prevalent.

Challenges

There are several reasons why sarcasm detection can be considered challenging:

- The sentiment/meaning of a sentence or paragraph can be inverted by one or more tokens (e.g., emoji, hashtag)
- Broader context of the overall statement is more important in evaluating whether it's sarcastic, and may be harder to detect via text compared to talking in person, as cues such as facial expressions and body language may not be available
- Sarcastic language often models serious language with sometimes subtle differences
- Hyperbolic language can be sarcastic, but not necessarily

Embeddings for tweets by tweets

Tweets are constrained by their character length, which challenges users to be creative in their post structure and content. The result is a fairly unique style of writing that can differ from the corpora used to generate the default embeddings for traditional large language models. Creating custom embeddings is important to align with the form and character of language in tweet to maximize model performance; however, customizing embeddings on a small-ish dataset may risk losing the generalizability of the default embeddings.

Datasets

The datasets derive from a collection used for the purposes of the following publication: UTNLP at SemEval-2022 Task 6: A Comparative Analysis of Sarcasm Detection using generative-based and mutation-based data augmentation. The datasets are described in the article and include:

- iSarcasm data
- Sarcasm Headlines Dataset
- Sarcasm Headlines Dataset v2.0
- SPIRS
- Twitter US Airline Sentiment
- Sentiment140 datasetmm
- Twitter News Dataset

The following serve as alternative and supplemental datasets:

- Baskin Engineering @ UC Santa Cruz: Sarcasm Corpus V2

Implementations

A range of small models from Huggingface will be used in combination with default and custom word embeddings. Preliminary models include: Distilbert, Albert and Bertweet

- # References

Article 1 - UTNLP at SemEval-2022 Task 6: A Comparative Analysis of Sarcasm Detection using generative-based and mutation-based data augmentation
Article 2 - Sarcasm Detection in a Disaster Context
Article 3 - Understanding the Sarcastic Nature of Emojis with SarcOji
Article 4 - BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding
Article 5 - Amrita_CEN at SemEval-2022 Task 6: A Machine Learning Approach for

Detecting Intended Sarcasm using Oversampling
Article 6 - Combining BERT with Static Word Embeddings for Categorizing Social Media