Analyzing Context and User Information in Online Sarcasm Detection

Mohan Dodda **Tusheet Sidharth Goli** **Aarun Srinivas**

College of Computing, Georgia Institute of Technology
{mdodda3, tgoli3, asrinivas5}@gatech.edu

Abstract

We aim at identifying sarcasm in conversations and examining the context and circumstances in which sarcasm tends to be expressed. By performing a sentiment, emotion, and clustering-based analysis on the data, we will categorize whether conversations sampled from diverse contexts are sarcastic or not. Based on this categorization, we aim at performing a deeper analysis of the features that warrant sarcasm: (1) users – demographic, background, popularity, social media indicators, etc., and (2) context – identify what topics/context are more likely to produce sarcastic conversations. We will investigate several types of Long Short-Term Memory (LSTM) or transformer (BERT) networks that can model both the conversation context and the sarcastic response for performing qualitative analysis. By analyzing the roles of users and the conversation context for sarcasm detection in online conversation (Ghosh et al., 2017), we intend on using this information for building a classifier that takes context/user information and generates if there will be a sarcastic response or not.

Introduction

Sarcasm is utilized in many informal and professional settings to express humor and cynicism as well as indicate hostility. Sarcasm also has a strong influence on the meaning and tone of a conversation. Therefore, due to its structural importance in modern-day text, the task of identifying and understanding the factors behind sarcasm is a useful application that can help aid in the construction of objective datasets. In particular, the datasets that can be formed from using the information obtained from our study will not only be more compact, but also have a much clearer meaning as sarcasm can often convolute or obscure the meaning of the text. This will allow models that will train on language data that have been neutralized by our approach to converge much faster as they will not have to accomplish the subtask of identifying sarcasm and understanding its meaning and impact on the surrounding text. Although this task may seem similar to identifying hate speech, it differs immensely due to the relatively low complexity associated with identifying hate speech. Specifically, when identifying hate speech, the solution is to just look for certain trigger words present in the sentence of interest. Sarcastic text, on the other hand, cannot be differentiated from the non-sarcastic text when examined in isolation. This means that context is important when identifying sarcastic text. Therefore, to identify sarcasm, more text than the sentence of interest needs to be fed into the classifier to produce an accurate identification. Furthermore, to understand which factors cause sarcasm, additional features not tied to the text, but rather tied to the user who generated the text need to be analyzed. Specifically, a user's demographic information, background, popularity, and personality all come into play when determining what causes a user to post sarcastic content.

Related Works

Compared to other natural language processing tasks, sarcasm detection and the causes behind sarcasm have been relatively unexplored. In fact, most research conducted on sarcasm detection has focused on determining whether a sentence is sarcastic or non-sarcastic in isolation (Davidov et al., 2010). In particular, this paper revolved around transforming a dataset containing sarcastic and non-sarcastic sentences into data points that consisted of hand-designed features, syntactic patterns, and lexical cues. Recently, however, there have been efforts like Wallace et al. (2014) in utilizing contextual features such as the author, topics, and conversational context of the text to aid classifiers in detecting sarcasm. These approaches also utilize hand-designed features like most of the isolated sarcasm detectors, but differ in that they also use embedding-based representation through deep learning. Other recent works also tackle the lack of context problem by attempting to understand the whole conversation surrounding the sentence of interest to detect sarcasm (Ghosh et al., 2017). Instead of hand-designing conversational context features, these works utilize several LSTMs to provide conversational context to their classifiers. Ghaeini et al. (2018) improves upon the work of Ghosh et al. (2017) further by looking at the sentence of interest in both isolation and in the context of the conversation. Finally, the most recent work that touches upon sarcasm detection utilizes pre-trained transformer models and recurrent convolutional neural networks with minimum feature engineering to outperform all other proposed methodologies (Potomias et al., 2020). Another direction we want to look at is which type of users produce sarcastic content which Ghaeini et al. (2018) does with user embeddings. To the best of our knowledge, there is no work that does a full-scale analysis on the concrete user and textual context that influence sarcasm. Therefore, our work will address this gap in research and contribute new knowledge by adding additional frameworks that will not only detect sarcasm, but also understand the factors that contribute to sarcasm.

Methods and Data

For this project, we will look at multiple datasets that include sarcasm with context. The first one will be Sarcasm Corpus V2. This dataset contains discussion forms, which have context (quotes to which the posts are replies) and their responses. These responses can either be sarcastic or non-sarcastic. The context in this dataset emulates real-world discussions and what exactly causes someone to use sarcasm. We will also consider Reddit. For this, we will utilize the SARC dataset which utilizes the /s tag to find sarcastic replies. This dataset also stores parent and child comments as contexts. We will explore different ways to extract context from the parent and child comments. Lastly, we will look at Twitter. To scrape sarcastic and non-sarcastic tweets. We will find a Bigquery Twitter Dataset and use specific tags such as #sarcasm and use the /s tone indicator to find sarcastic tweets with the "reply to status" parameter to get the replies to these tweets. Additionally, we want to mine the information about the users who produce sarcastic content. For this, we will mine the user information associated with a sarcastic tweet. We will look at the number of followers, the location they are from. For data validation, we will run a pre-trained sarcastic BERT model on it to weed out the non-sarcastic responses.

With this data, we will look at different models to try to get an understanding of the context of sarcasm. Using a BERT model (or using variants), we will finetune it on our dataset to make a better sarcasm classifier. We will use the labeled replies for this task without any context. This will give us a baseline predicting sarcasm. To see if we can predict if a particular context will result in sarcasm, we will fine-tune another BERT model to see if we can predict if a sarcastic response will occur solely from the replies. Additionally, we will try to extract features from the context that may prompt a sarcastic reply. We will do this by clustering the non-sarcastic and sarcastic contexts with LDA. We will also do clustering with user information from user embeddings from CASCADE and user information from Twitter. We will also perform sentiment analysis. We can also dive deep into the dataset and manually find trends and categories that the sarcastic context fits into. With our sarcastic classifier, we can also give a score to the sarcastic replies and see which types of context influences more sarcastic replies. We can do this analysis on all three datasets as in all of these datasets, we have sarcastic/non-sarcastic replies and context. Lastly, we will expand upon CASCADE (Hazarika et al., 2018) to utilize a BERT classifier instead of an LSTM variant to see if we can improve results in predicting sarcasm.

Timeline

We plan on splitting this entire project into 4 phases. We plan on dedicating phase 1 (Feb Week 1 – end of Feb Week 4) for data collection and cleaning. Since we are looking at 3 different datasets, Sarcasm Corpus V2 [Mohan], Bigquery Twitter Dataset [Aarun], Reddit Sarcasm Dataset (SARC) [Tusheet]; each member of the team will be responsible for one dataset and cleaning it. In phase 2 (March Week 5 – mid-March Week 7), we firstly want to replicate the LSTM model (Ghosh et al., 2017) and the CASCADE model (Hazarika, 2018) on our data [Mohan] while implementing the clustering and LDA based analysis for categorizing sarcastic conversation [Tusheet and Aarun]. Following this in phase 3 (mid-March Week 8 – early April Week 10), we will then train our classifiers (BERT and LSTM) on the context as well as embedded user data to predict sarcastic conversations. While Mohan works on embedding user information and context to the data, Tusheet and Aarun will work on the classifiers that are trained to predict sarcasm in conversations based on context/user data. In the fourth and final phase (early April Week 11 – deadline Week 14), we will work on analyzing the final results, plots, and work on writing the final deliverable paper and the presentation. We will set a deadline to complete all these deliverables at least one week before their due date. This will be done collectively by all members of the team.

References

Dmitry Davidov, Oren Tsur, and Ari Rappoport. 2010. Semi-supervised recognition of sarcasm in twitter and amazon. In Proceeding of CoNLL, 2010., pages

107–116.

Reza Ghaeini, Xiaoli Z. Fern, and Prasad Tadepalli. 2018. Attentional multi-reading sarcasm detection. CoRR, abs/1809.03051.

Debanjan Ghosh, Alexander Richard Fabbri, and Smaranda Muresan. 2017. The role of conversation context for sarcasm detection in online interactions. In Proceeding of SIGdial Meeting on Discourse and Dialogue, 2017., pages 186–196.

Potamias, R.A., Siolas, G. & Stafylopatis, A . A transformer-based approach to irony and sarcasm detection. Neural Comput & Applic 32, 17309–17320 (2020). https://doi.org/10.1007/s00521-020-05102-3

Byron C. Wallace, Do Kook Choe, Laura Kertz, and Eugene Charniak. 2014. Humans require context to infer ironic intent (so computers probably do, too). In Proceeding of ACL, 2014., pages 512–516.