# Analyzing Context and User Information in Online Sarcasm Detection

**Mohan Dodda** **Tusheet Sidharth Goli**

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

**Abstract**

We aim at identifying sarcasm in conversations and examining the context and circumstances which provoke sarcastic responses. Using the SARC Reddit Dataset, we have improved upon baseline Logistic Regression models, utilizing BERT based models in sarcasm detection. We proposed and showed new results in predicting sarcasm from parent comments and a potential sarcastic response. Lastly, we analyzed features that exist in sarcastic responses including punctuation, subjectivity, profanity, and comment length and their correlation with sarcasm. In the future, we will utilize user embeddings and expand up on the CRAFT model with BERT: CRAFT+BERT. We will mine for different features in parent comments that provoke sarcastic results. For this, we will look at new features such as sentiment analysis, emotion. In the user side, we will mine different user related features such as popularity, user background, and user interests. We will also cluster pretrained user embeddings and discover new features. Lastly, with all of the new user and contextual based features, we will make a model utilizing only these features to determine which combination of features work well in detecting sarcasm the best.

## Introduction

Sarcasm is utilized in many informal and professional ways to express humor and cynicism as well as indicate hostility. Sarcasm strongly influences the meaning and tone of a conversation. Therefore, sarcasm is an important aspect of human language and psychology as it is prevalent in all cultures. However, detecting sarcasm is difficult, even to humans without the correct context. This task is important as it can help models converge faster as they do not have to accomplish the subtask distinguishing sarcastic responses to truthful responses. Without this distinction, it would be easy to confuse sarcasm with incorrect and out of place responses.

In this context, we look at sarcasm in online settings as those are the datasets we have available. Sarcasm, in online settings, basically contains the comment(s) and the sarcastic response to the comment. This shows that context is necessary for sarcasm and that sarcasm doesn't come from nowhere. What type of contexts provoke sarcasm? What causes this sarcasm? Understanding what can cause sarcasm can help us understand human psychology and human language at a much deeper level which can allow us to create better language systems and more objective datasets. Furthermore, sometimes, the parent comment to a sarcastic response is not only what determines sarcasm. Certain people are more likely to produce sarcastic comments. Factors such as background, popularity and personality all determine if a person responds with a sarcastic quip. By doing user based analysis, we can see what type of people are more likely to produce sarcastic responses as well.

Let's say a person always gets a lot of sarcastic responses to their comments, understanding what type of comments they are saying to provoke these sarcastic responses can help them reduce these comments and who they are talking to will help as well. This is important because sarcastic responses are not always desired and analyzing the contexts that cause the sarcasm can prevent tension. Also, for some people it is hard for them to distinguish if a response is sarcastic or not as they are not knowledgeable about sarcasm. This can cause a lot of confusion and even the feeling of being left out. By understanding the types of comments they are saying and who they are talking to, they can learn more about sarcasm and understand which speaking pattern of themselves causes sarcasm.

The goal of the project has not changed much. It has changed in that we are focused a bit more on identifying sarcasm itself more than originally intended. We did this more in this midway report. This is because we felt that identifying sarcasm is important to understand what factors cause sarcasm. We also feel that we can improve upon baseline results in this area as well.

## 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) improve upon the work of Ghosh et al. (2017) further by looking at the sentence of interest in both isolations and in the context of the conversation.

In order to accumulate accurate predictors for sarcasm, it is essential to delve deeper into the context, language, and user attributes that have high correlations with sarcasm. These reliable indicators will help our model to better predict sarcasm by identifying these specific features. Studies show that profanity in language is a great indicator and instigator for sarcastic behavior (Du et. al. 2021). Not only is there a strong correlation between sarcasm and profanity, but instances of profanity in parent comments is one of the major instigators for sarcasm. Research has also shown that sarcasm is often displayed in contexts that involve subjective conversations (Voyer et. al. 2010) where people are sharing their personal ideas and opinions, rather than basing their arguments on facts (Saha et. al. 2017). Other linguistic cues like intensifiers, capitalization, word unigrams, etc. have also been shown to be strong features of sarcasm (Bamman et. al., 2015). Taking inspiration from these methods and features, we have used these strong coreallational features in addition to researching a few more language and context based attributes to effectively predict the probability occurrence of sarcasm.

Similar research has showcased prominent user attributes and features that are a great predictor for sarcastic behaviors. Features such as user online personality, location, followers, demographics, etc. can play a crucial role in the type of online presence a user exhibits (Lynn et. al. 2019). These context, language and user based features can be embedded into a deep neural network architecture which can be used to predict sarcastic behaviors based on these cues (Amir et. al., 2016). These neural network architecture based context and user embeddings have proven to be an accurate estimator for sarcasm (Kumar et. al. 2018). 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) do with user embeddings. Hazarika et. al. (2018) also utilizes user embeddings but also uses a CNN network with the SARC reddit dataset (Khodak et. al. 2018). 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.

## Data

For this project, we originally planned on looking at multiple datasets. These include the benchmark Sarcasm Corpus V2, the SARC dataset from Reddit, and scraped Reddit dataset. However, from feedback from the instructor regarding the difficulty of getting permission to utilize their API keys, we decided to not go through the process to get permission to scrape Twitter even though Ghaeini already had data available. This is because The SARC dataset already has context and user information itself! Now looking at the SARC Reddit dataset, there are two separate sources we are looking at. The other is from A Large Self-Annotated Corpus for Sarcasm (Khodak et. al 2018) which mainly focuses on the textual part including the parent comment to the sarcastic comment as the “context”. The other is from CASCADE: Contextual Sarcasm Detection in Online Discussion Forums (Hazarika et al. 2018) . This also includes user embeddings to perform user-based clustering. For the midway point, we just focused on the textual context analysis so we just looked at the first source for now. We simply collected this dataset from the Kaggle competition associated with this dataset. (https://www.kaggle.com/danofer/sarcasm) . This dataset is quite large having 1.3 million Sarcastic comments. Additionally, the dataset has other related information including upvotes, downvotes, when it is

created, and the subreddit. We utilized the balanced version of the dataset which has an equal amount of sarcastic and non-sarcastic responses.

| Statistics vs. comment type | Sarcastic | Non-Sarcastic |
|---|---|---|
| average response size (words) | 10.33 | 10.59 |
| average parent comment size | 24.21 | 24.56 |
| mean upvotes | 5.22 | 5.78 |
| mean downvotes | -0.13 | -0.17 |
| mean score | 6.40 | 7.37 |
| Number of Documents | 505368 | 505405 |
| label | 1 | 0 |

Table 1: SARC Dataset Statistics over the label

Regarding preprocessing, we get rid of rows with any null values in the comment or the parent comment section. For the logistic regression and the clustering, we removed stop words. We also do a 75-25 train test split for our models.

**Here are examples of a Sarcastic data point:**

**Parent Comment:** "Those statistics on Reddit are typically overstated"

**Sarcastic Comment**: "Got a source on that?"

**Explanation:** The sarcastic comment is a quippy response to the fact statistics are overstated without sources. The comment speaks to the parent comment asking for the source for the statistic about statistics.

**Parent Comment:** "Armed robbers stormed a luxury hotel in central Berlin where a poker tournament was taking place, German police say."

**Sarcastic Comment**: " 'Obviously they need to ban video games so it can never happen again."

**Explanation:** This speaks to the very common misconception that video games cause violence where in this case the robbery was happening where non-video game was taking place.

**Example of a Non-Sarcastic data point:**

**Parent Comment: "**What is NOT a fun fact?"

**Non-Sarcastic Comment:** "Prion diseases are always fatal"

**Explanation:** This response answers the question correctly and thus isn't sarcastic. If the answer was incorrect, then there would be a chance the response is sarcastic.

Now for Sarcasm Corpus V2, we realized that we need to request access through an online form. We requested this a while ago with no response yet. Therefore, for this midway report, all the analyses will be utilizing the SARC Reddit dataset. We might decide to drop the usage of Sarcasm V2 as we realized the SARC dataset is already really big and diverse with user information as well! That is why we do not perform any analysis of Sarcasm Corpus V2 here. We decided to save the user-based analysis for the final section. Therefore, if we decide that the user information and embeddings from SARC are insufficient, we will explore Ghaeini's Twitter dataset.

## Method

Within this data, we are both analyzing results and developing models. In the final report, we will focus much more on context analysis. Thus, for the midway report, we focused much more on the sarcastic responses. First, we run models on predicting sarcasm. For this, we consider 3 x values: the sarcastic comment only, the parent comment only which acts as the context, and both the parent comment and the sarcastic comment. For the latter one, we concatenate the parent comment and the sarcastic comment. The purpose of this is to compare the role of context in sarcasm. We also want to see if it is feasible to just use context to predict if it will result in a sarcastic response. We utilize two models. The first is the baseline utilized in Khodak et. al. (2018): a logistic regression. We use this as a baseline. For the next model, we utilize a new model not used commonly in BERT. We finetune a BERT model on each of the x values to predict sarcasm. In the future, we want to extend even further and utilize the CASCADE model. This utilizes the user embeddings and CNN-based model. We look to extend this and make it BERT-based rather than CNN-based. Lastly, we will utilize the full context, all of the text before the potential sarcastic

comment, not just the direct parent, from Khodak et. al. (2018). On the other side, we are trying to analyze factors and topics that cause sarcasm. For this, we look at clustering methods. For now, we focus on the sarcastic responses for now.

In addition to this, we worked on identifying certain context and language-related features that tend to correlate with sarcastic comments. We analyzed this in two main aspects - parent comment and sub-comments. Our intuition behind this was that there might be certain features that are a part of the parent comment (main post) that could encourage/instigate sarcastic replies. In addition to this, we also wanted to characterize features that correlate with sarcastic messages. We nailed down factuality (subjectivity), score, upvotes, downvotes, average word length, average sentence length, special characters (?!), profanity score, parts-of-speech (verb, adjective, pronoun, noun), and overall context analysis for the parent and the sub-comments to identify any correlating factors that tend to promote sarcastic behavior. We used python's TextBlob NLP module to perform our sentiment parts-of-speech analysis. Our approach was to analyze important aspects of our data that we thought would encourage this type of behavior. We also read a lot of literature and papers to understand some of the factors that previous research had found useful. We thus used our intuition as well as related works in this field to nail down some of the key attributes we wanted to analyze in detail.

## Preliminary Results

We have implemented preliminary models on the preliminary dataset. This includes the Logistic Regression model and the BERT model on the parent comment only, potential sarcastic comment response only, and parent plus response. We still want to utilize the whole parent thread and also utilize a CRAFT+BERT model. These preliminary models are successful in addressing one of the project goals which is improving upon baseline models which predict sarcasm. Also, it explores just utilizing the parent comment in predicting sarcasm and if it is feasible to utilize just the parent comment to predict sarcasm. Our success metric is Accuracy. We do not utilize F1 scores now as our data is balanced. However, we will still utilize it later. As stated before, the dataset is split into a 75-25 test split where the model is trained on the train set and accuracy is calculated on the test set. For the Logistic Regression model, we use the liblinear solver with a C of 1.0. For the BERT model, we train 1 epoch utilizing the BERT-base. We use a batch size of 24 and a max_length of 128. Here we cut the length of each data point to a max length of 128. We use AdamW optimizer with a gradient clipping.

| Dataset vs model | Logistic Regression | BERT - full dataset |
|---|---|---|
| Only Response | 0.72 | 0.78 |
| Only Parent Comment | 0.58 | 0.61 |
| Parent Comment and Response | 0.68 | 0.74 |

Table 2: Table Sarcasm Accuracy Comparison

We match the baseline in Khodak et. al. (2018) with our logistic regression implementation getting .722 accuracy. Across all of the datasets, the BERT model gets better results than the Logistic Regression model. The BERT model is better able to capture the semantic differences between the texts. Additionally, the BERT performance of .78 is better than using CRAFT which gives us .77. CRAFT uses a CNN and has additional information from user embeddings. This shows the power of self-attention and further reinforces the future work of exploring a BERT+CRAFT model. The only Parent Comment results make sense giving lower results than using the sarcastic response. This is because this task is much harder. We are seeing if a particular comment would prompt a sarcastic response and not all parent comments would prompt sarcastic comments. Additionally, there are a lot of parent comments that prompt both sarcastic and non-sarcastic comments. Therefore a .61 accuracy is still impressive showing it would be useful to analyze features that "provoke" sarcasm. Lastly, the results that include both the parent comment and response as the train data are not what was expected. This concatenation of the parent comment with the potential sarcastic response gives us a lower score than just the response which is the opposite of what we were expecting. This is probably because the parent comment is very noisy. A better approach in combining context with the sarcastic comment might give us better results. We did not explore different hyperparameter values. This is because we wanted the Logistic Regression to act as a baseline. For the BERT models, we did not have the time/resources to train with different hyperparameters, but will explore it in the final report!

We also performed an analysis on context and language-related features that could correlate to sarcastic behaviors. We analyzed the following attributes for

sarcastic and non-sarcastic messages for both parent comments and their relevant sub-comments:

Average word length - This attribute did not showcase any strong correlation between sarcastic and non-sarcastic texts. This aspect is generally not a good indicator of sarcasm and any correlations from this aspect are mostly coincidental or highly data specific.

| Sarcastic | Non-sarcastic |
|---|---|
| 4.379 | 4.278 |

Table 3: Average number of letters in words

Average sentence length - Much like the word length, this attribute did not showcase any strong correlation between sarcastic and non-sarcastic texts. And similar to average word count, this aspect is generally not a good indicator of sarcasm and any correlations from this aspect are mostly coincidental or highly data specific.

| Sarcastic | Non-sarcastic |
|---|---|
| 40.451 | 40.813 |

Table 4: Average Number of words in a sentence

Count of special characters (?!) - There showed a strong correlation for sarcastic text containing the '!' character than non-sarcastic text. This shows that '!' marks could be used as a good predictor for sarcasm. Unfortunately, this did not transpose over to the '?' character as it showed minimal correlation.

Special Character Comparison

| Character | Sarcastic | Non-sarcastic |
|---|---|---|
| ! | 6884 | 2727 |
| ? | 5564 | 6306 |

Table 5: Amount of times a specific special character appears for sarcastic and non-sarcastic texts

Subjectivity - We analyzed the subjectivity behind both the parent comments as well as its corresponding sub-comments. This basically corresponds with how factually correct a response is. Although the parent comment did not show any trends, the sub-comments have a higher degree of overall subjectivity, i.e., sarcasm is more prevalent when the context is antithetical rather than being factual. Our results are in line with prior research between sarcasm and subjectivity (Saha et. al., 2017). This indicates that sarcasm tends to be more prevalent in situations when people are sharing their own personal opinions rather than basing their arguments on facts (Voyer et. al. 2010).

| Type | Sarcastic | Non-sarcastic |
|---|---|---|
| Parent comment | 18625.377 | 18114.239 |
| Sub-comment | 16510.655 | 15225.279 |

Table 6: Subjectivity score of dataset partitions with higher score representing a higher subjectivity

Profanity count - We counted the number of profane words in these comments and found that sarcastic comments are generally associated with a larger degree of profanity and vulgarity compared to non-sarcastic comments. Furthermore, parent comments that have a larger degree of profanity tend to provoke a greater amount of sarcastic replies. Our results are in line with prior research between sarcasm and profanity (Du et. al.). This feature can be used as a decent predictor for sarcasm detection.

| Sarcastic | Non-sarcastic |
|---|---|
| 5766 | 5333 |

Table 6: Number of Profane words by the label of the dataset

Lastly, we analyzed word clouds among the top 5 sub-Reddits in this dataset to find any common underlying themes within the sarcastic replies. It was noticeably observable that most of the comments that were sarcastic indicated nationality, demographic, religious, gender, or other stereotyping tones in their replies. Irrespective of the context, these features in reply texts showed a significant amount in these conversations. Although our results for this analysis were not too quantitative, these strong correlations showcased by the word clouds on stereotyping behaviors based on nationality, gender, religion, etc. could potentially be used as a great indicator to identify sarcastic responses. Having successfully identified some of the potential indicators for sarcasm, we plan on using some of these indicators for the later phases of our project.

## Timeline

We propose a four-phase approach to this project. We plan on dedicating phase 1 (Feb Week 1 – end of Feb Week 4) for data collection and cleaning. Since we are looking at 2 different datasets, Sarcasm Corpus V2 [Tusheet] and Reddit Sarcasm Dataset (SARC) [Mohan]; each member of the team will be responsible for one dataset and cleaning it. This phase of the project has been successfully completed. Mohan worked on collecting and cleaning the Reddit dataset as well as worked on the Logistic Regression and BERT model for the parent comment. Tusheet worked on individual context and linguistic feature analysis as well as other clustering methods for the topics in the sub-comments. Both members equally contributed to all aspects of the midterm report and the presentation. 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 et al. 2018) on our data [Mohan] while implementing the clustering and LDA based analysis for categorizing sarcastic conversation [Tusheet]. 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 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.

## Future Tasks

Based on our 4 phase plan, we have successfully completed the first phase of the data collection and cleaning process. On top of that, we were able to complete our preliminary classification of sarcasm (sarcastic/non-sarcastic) and were able to successfully identify some key attributes/features that promote sarcastic behaviors online. These features are essential to the process of analyzing content attributes in sarcasm in the future phases. We plan on building upon these features/attributes that encompass the first half of our hypothesis testing, i.e., context and language-related features in sarcasm. This is the short-term plan at hand.

We are sticking with our initial plan of action as we have successfully planned the scope of the project until now and have made meaningful progress in order to achieve our main goal for this project. Moving forth, we are planning on specifically analyzing user information-related features, much similar to how we analyzed the context and language-related features. We have decided to use the SARC Reddit dataset to analyze particular user attributes and traits that could potentially help us to draw a correlation with sarcasm. We plan on using the LSTM model (Ghosh et al. 2017) and the CASCADE model (Hazarika et. al. 2018) presented in these papers to perform this analysis. Based on these findings, after we have finalized the particular features we want to analyze, we will train our classifiers by embedding the context and user features. These classifiers will include an extension on CASCADE using BERT: BERT+CASCADE other BERT variants including BART and DistilBERT. These classifiers will be used to predict sarcastic comments given some context, language, and user-related features. Analyzing these results and potentially getting positive results based on our feature hypothesis will have confirmed our hypothesis. We plan on doing intensive feature analysis to ensure these features have a prominent correlation with sarcastic comments. With this methodology at hand, we expect to get positive prediction results for sarcasm given these features. Lastly, we will create a feature based model to predict sarcasm. These features will only be the user and parent comment based features. We will utilize an interpretable machine learning model XGBoost or Logistic Regression and discover which combination of features work the best in predicting sarcasm. This is the bigger picture/long-term plan for our project. The four-phase plan and the individual split of responsibilities are detailly outlined in the timeline section of our report.

## References


Amir, S., Wallace, B., Lyu, H., and Silva, P. 2016. Modelling context ´ with user embeddings for sarcasm detection in social media. arXiv preprint arXiv:1607.00976.

Bamman, D., and Smith, N.. 2015. Contextualized sarcasm detection on twitter. In Ninth International AAAI Conference on Web and Social Media.

Davidov, D., Tsur, O., and Rappoport, A. 2010. Semi-supervised recognition of sarcasm in twitter and amazon. In Proceeding of CoNLL, 2010., pages 107–116.

Du, H., Dong, W., Lopez. K. 2021. Detecting the Emotion Dynamics of Profanity Language In Social Media Through Time

Ghaeini, R., Fern, X., and Tadepalli, P. 2018. Attentional multi-reading sarcasm detection. CoRR, abs/1809.03051.

Ghosh, D., Fabbri, A., and Muresan, S. 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.

Hazarika, D., Poria, S., Gorantla, S., Cambria, E., Zimmermann, R., and Mihalcea, R. 2018. CASCADE: Contextual sarcasm detection in online discussion forums. In Proceedings of the 27th

International Conference on Computational Linguistics, pages 1837— 1848. Association for Computational Linguistics.

Khodak, M., Saunshi, N., Vodrahalli, K. A large self-annotated corpus for sarcasm. In Proceedings of the Linguistic Resource and Evaluation Conference (LREC), 2018.

Kumar, L., Somani, A., & Bhattacharyya, P. 2017. Approaches for computational sarcasm detection: A survey. *ACM CSUR*.

Lynn, V., Giorgi, S., Balasubramanian, N., & Schwartz, H. A. 2019. Tweet classification without the tweet: An empirical examination of user versus document attributes. In Proceedings of the third workshop on natural language processing and computational social science (pp. 18–28). Association for Computational Linguistics, Minneapolis, Minnesota, http://dx.doi.org/10.18653/v1/W19-2103.

Potamias, R.A., Siolas, G. and 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

Saha S., Yadav J., Ranjan P. 2017. Proposed approach for sarcasm detection in twitter. Indian J Sci Technol 8

Wallace, D., Choe, D., Kertz, L., and Charniak, E. 2014. Humans require context to infer ironic intent (so computers probably do, too). In Proceeding of ACL, 2014., pages 512–516.