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

This paper delves deep into identifying sarcasm in conversations and examining the context and circumstances which provoke sarcastic responses. This is done by examining context and circumstances that provoke sarcastic responses as well as analyzing linguistic cues and user attributes that correlate to high sarcastic behaviors. 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. We analyzed features that exist in sarcastic responses including punctuation, subjectivity, profanity, and comment length and their correlation with sarcasm. We further analyzed user attributes by utilizing user embeddings and expanded upon the CASCADE model with BERT: BERT+CASCADE. We mined for different features in parent comments that provoke sarcastic results. We looked at new features such as sentiment analysis, profanity, subjectivity, special characters, and emotion. On the user side, we mined different user-related features such as popularity, user background, history, and user interests to create an overall user personality metric. Lastly, with all of the new user and contextual-based features, we made a model utilizing these features to determine which features are most important to predict sarcasm. Our work can help people distinguish between incorrect information and sarcasm in online platforms as well as provide a deeper understanding of human psychology and help in linguistic analysis.
GitHub - https://github.gatech.edu/mdodda3/CSS_Sarcasm

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 for 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 of distinguishing sarcastic responses from 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, especially on online social media platforms. Sarcasm, in online settings, basically contains the parent comment(s) and the sarcastic response to the comment. This implies the importance and necessity of context in sarcasm and that sarcasm doesn't come from nowhere (Wallace et. al., 2014). The questions we want to answer are: What type of contexts provoke sarcasm? What causes this sarcasm? Are there any common correlational features that provoke 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. As prior research has shown, factors such as background, popularity, and personality all determine if a person responds with a sarcastic quip (Amir et. al., 2016). 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 (Batista et. al., 2022). 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. These negative connotations associated with sarcastic replies have been a part of multilingual studies conducted by various researchers in the past and present (Syafruddin et. al., 2021). 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. Our project goals have not changed much and are consistent with the original goals that were proposed in the proposal and midway report.

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 the language is a great indicator and instigator of sarcastic behavior (Du et. al. 2021). Not only is there a strong correlation between sarcasm and profanity, but instances of profanity in parent comments are one of the major instigators of 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 correlational 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 of sarcastic behaviors. Features such as using 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 contexts, 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 accurate estimators 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 utilize user embeddings but also use 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 of 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 instructors, we decided to only use the SARC (Self-Annotated Reddit Corpus) Reddit dataset since it already has context and user information itself and is a popular dataset for online sarcasm detection and analysis (Khodak et. al., 2018). Our data has been consistent with our proposal and midway report and has not changed from the midway report. Now looking at the SARC Reddit dataset, there are two separate sources we are looking at. The first 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.

We 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 parent and sub-comments that may or may not be associated with a sarcastic response from a wide range of subreddits. 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. Due to its large size and prevalence in sarcasm studies, we can attribute the generality of the study. We performed a deep contextual analysis of several aspects of this dataset to extract features that correlate to higher sarcastic behaviors.

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 a 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.

Methods

Within this data, we are both analyzing results and developing models. For the midway report, we focused much more on the sarcastic responses and context feature analysis for sarcasm. In the final report, we shifted our focus much more towards user-related features of sarcasm and implementing the BERT+CASCADE model. 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.

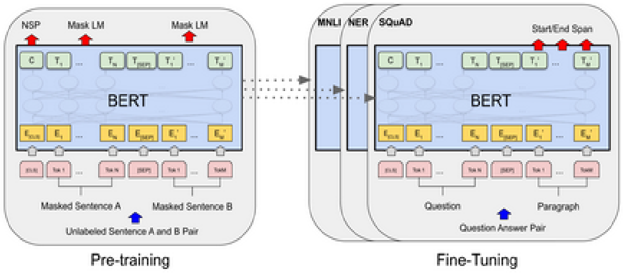


Figure 1: BERT model fine-tuning process

Lastly, we utilize the CASCADE model. This utilizes the user embeddings and CNN-based model. The user embeddings are created by generating user personality and user style by reading from all of the text the user has generated. We extend this and make it BERT-based rather than CNN-based to create a BERT+CASCADE model. We first generate user embeddings in the same way indicated in CASCADE. Now the model is constructed as such: We feed in the text through the BERT as normal. However, the output from BERT is then concatenated with the User Embedding into a linear layer and an output layer.

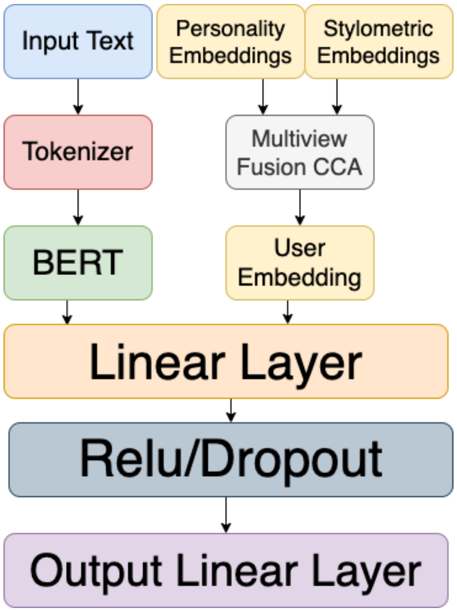


Figure 2: BERT+CASCADE Model Architecture

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.

Utilizing these features of high correlation value to sarcasm, we developed a feature model. This predicts sarcasm by using these contextual and user-related features; however, the main goal for this is to see which features are most impactful in predicting sarcasm! For context-related features, we incorporated sentiment (positive/negative connotation), subjectivity (degree of factuality), profanity (degree of vulgarity), and count of special characters (!?). We utilize TextBlob to calculate a text's subjectivity and sentiment which outputs a score of [0,1] for subjectivity and [-1,1] for polarity. Profanity is a binary label on if a piece of text contains profanity or not. For the user-related features, we extracted the author name, subreddit, and author personality from the BERT+CASCADE model. For the user personality, we extracted user embeddings from the BERT+CASCADE model and utilized PCA to reduce a 150-dim vector to a 1-dim vector. We do this in order to capture the user's personality in one interpretable dimension. If we reduced it to let's say two dimensions, we would not be able to distinguish between the two user personality dimensions. We thus created a concatenated 7-dim context and user feature vector and trained an XGBoost model to predict sarcasm based on these features. and also identify the context and user-related features that were the most significant at identifying and predicting sarcasm.

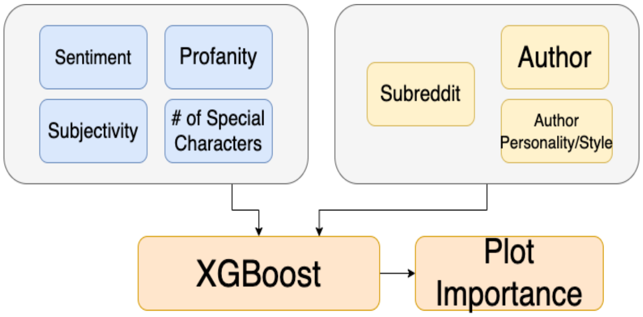


Figure 3: Feature-based Prediction Model

Results

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 CASCADE which gives us .77. CASCADE 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+CASCADE 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.

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 2: 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 3: Average Number of words in a sentence

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

Special Character Comparison

Character	Sarcastic	Non-sarcastic
!	0.72	0.28
?	0.47	0.53

Table 4: 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 their 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 factual. Our results are in line with prior research on 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	0.51	0.49
Sub-comment	0.52	0.48

Table 5: Subjectivity score of dataset partitions with a 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
0.52	0.48

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

Lastly, we analyzed word clouds among the top 5 subreddits 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, religion, 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 for identifying 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.

Experimental Setup

We have two areas we look at: sarcasm detection utilizing text and sarcasm detection using only features. For the text-based results, we construct three separate text-based datasets: the parent comment, the response comment, and the parent and response comment concatenated. We run all of our models on these three datasets. For the text-based sarcasm detection, we utilize the full balanced dataset of size 1010736. We utilize a train/test split of 75-25. Our success metric is Accuracy. We do not utilize F1 scores now as our data is balanced. As mentioned above, we train our datasets on three different models: Logistic Regression baseline, BERT-base, and BERT+CASCADE.

For our logistic regression model, we first utilize a TFIDF-Vectorizer to convert the text data into a numeric format. We do not have stop words and we utilize 1 and 2 ngrams in our dataset. Afterward, we utilize a standard Logistic Regression Model with C=1.0, so a liblinear solver. For our BERT and BERT+CASCADE model, we utilize a 16GB NVIDIA TESLA P100 GPU provided on Kaggle. We train for 1 epoch (we did not find improvements with more epochs as the dataset is already very big) with a batch size of 24. We cut the length of each data point to a max length of 128. It takes around 3 hours for both the BERT and BERT+CASCADE models to run on all the text datasets. We run our models utilizing AdamW optimizer, with a 2e-5 learning rate and epsilon 1e-8. We utilize a learning rate scheduler which reduces the learning rate to 0 after a warm-up period. We utilize cross_entropy loss and gradient clipping to prevent exploding gradients. No annotations were needed for this project.

For the feature-based model, we utilize an XGBoost model with learning rate=0.1, n_estimators=100 and max_depth=3.

Result Comparison

We run our models instructed in the Experimental Setup and get results dedicated to Chart 2. First, we replicate the Logistic Regression done by Khodak (2018) who created the Dataset. We do the same with BERT, BERT+CASCADE, and the other baseline CASCADE. We find that the neural network models (BERT, BERT+CASCADE, and CASCADE) all outperform the Logistic Regression. This can be explained by the power of neural networks in capturing text information better. This also displayed the negatives in the bag of words representations to capture lexical information such as sentence context. Lastly, we have so much data so neural networks are able to take advantage of that.

Also, within these comparisons, we see that the transformer-based models are much better at capturing Natural Language with the utilization of transformers and vast pre-training. This shows that BERT is very good for capturing nuances in sarcasm. The BERT performance of 0.78 is better than using CASCADE which gives us 0.77. CASCADE uses a CNN and has additional information from user embeddings. This especially shows the power of self-attention as it outperformed the addition of user embeddings into the model. Lastly, we look at the BERT+CASCADE model. We find produced results of .8 this is an improvement on the BERT of .78. (Notice that the result from our model is different from the presentation as we realized a bug within the training process and retrained to get .8 this time) Here we produce a novel result that has increased upon previous baselines. This is an improvement on both the BERT+CASCADE and BERT models. This demonstrates the power of utilizing external context in predicting sarcasm.

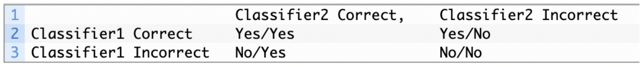
Next, we also train the models on the concatenated Parent with its response to explore the role of context in predicting sarcasm. We also just see if we can predict sarcasm with only the parent comment - with no expectation of great results.

By using both parent and response to predict sarcasm, We get a similar progression from the only response dataset where the BERT model of .74 is greater than the training logistic regression of .68 and the BERT+CASCADE model is greater than the BERT model. Similar conclusions made from the only response models can be concluded from these models as the user and context features show to improve upon regular BERT results. However, across the board, this dataset that adds the parent comment produces worse results compared to just the response.

Our original hypothesis was that utilizing a parent comment in addition to utilizing a response would only help as the parent comment produces contextual information as to why a sarcastic response was formulated. However, we easily see that this is not the case. This might be due to the noise in the parent comment. As indicated in our dataset analysis, we found that the parent comment is much larger than the potential sarcastic quip to it. The parent comment is probably too noisy to learn to use in conjunction with the regular comment. To fix this, we probably need to look at different ways to combine the parent and response comments. One way could be to make separate BERT models for both and concatenate the outputs alongside the user embeddings during the training process before feeding them into linear layers. Maybe, we need to capture the parent comment into a lower-dimensional feature space before we feed it into our model.

Lastly, we run our models on just the parent response. As expected, the results are not great and are considerably lower than the other two which utilize the response. Parent comments are not sole indicators of predicting sarcasm. It does follow the same forward progression as the other two models from .58, .61, to .60. This does not follow a similar progression as the BERT+CASCADE is slightly lower than BERT. Nevertheless, the results are not 50%. Those accuracies are still impressive. This demonstrates the power of BERT and its ability to extract what language people use from the language used to provoke sarcasm. One problem might be that CASCADE extracts user info from the response comment and maybe one that extracts from the user info of the parent comment would be useful to do instead. In this case, the user embeddings most likely acted as noise. Nevertheless, this shows promise that analyzing features that provoke sarcasm can be useful.

We also perform a McNemar's Test (Dietterich 1998) significance test over our models. We look at two different models (BERT+CASCADE and Logistic Regression) and two different datasets (Only Comment, Only Parent comment). The McNemar's test can be used to compare algorithms with a binary label without resampling methods, having to run our model multiple times. McNemar's test operates upon a contingency table. Now for each of the two models, we see if the respective model is wrong or right for every data point. We create a chart summing up the matches/mismatches of the incorrect and correctness of the models as displayed in figure 4.



```
                      Classifier2 Correct,   Classifier2 Incorrect
Classifier1 Correct   Yes/Yes                Yes/No
Classifier1 Incorrect No/Yes                 No/No
```

Figure 4: Table Sarcasm Accuracy Comparison

Then we calculate a statistic = (Yes/No - No/Yes)^2 / (Yes/No + No/Yes). The test statistic has a Chi-Squared distribution with 1 degree of freedom. Now if our statistic has p<.05, we can say that the improvement of the BERT+CASCADE was significant.

Only Parent Comment (Percentage)	BERT + CASCADE Correct	BERT + CASCADE Incorrect
Logistic Regression Correct	0.466	0.119
Logistic Regression Incorrect	0.137	0.277

Table 7: Parent Comment Table Analysis

Only Comment (Percentage)	BERT + CASCADE Correct	BERT + CASCADE Incorrect
Logistic Regression Correct	0.595	0.0749
Logistic Regression Incorrect	0.204	0.126

Table 8: Parent Comment Table Analysis

Now looking at both of our results, we can see that the results are indeed significant. For the only parent comment, the dataset BERT+CASCADE model's improvement was definitely statistically significant giving a p-value of 0 from a statistic of 27580. Similar results exist for just the response comment giving us a p-value of 0.0 and a statistic of 4845. This just shows how drastically the BERT+CASCADE model improved upon the results.

Dataset vs model	Logistic Regression	BERT	BERT + CASCADE
Only Response	0.72	0.78	**.81**
Only Parent Comment	0.58	0.61	.60
Parent Comment and Response	0.68	0.74	.79

Table 9: Table Sarcasm Accuracy Comparison

Next, we do some error modeling and we see where our best model: BERT+CASCADE performed badly looking at a confusion matrix in figure 5. We see that we have around 2.5 times more false positives than false negatives. This means that when the model is not good at predicting non-sarcastic results and tends to be more optimistic, saying things are sarcastic most times. This might be because sarcasm can be really subjective and really close to just a false statement. Thus if a model finds a statement that seems to be incorrect it just labels it as sarcastic.

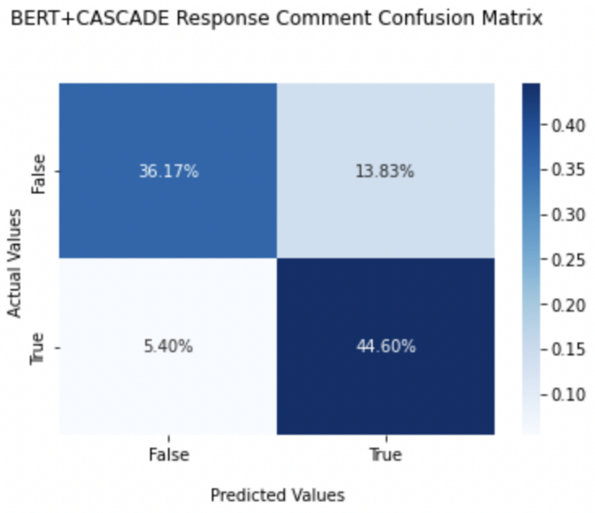


Figure 5: BERT+CASCADE Error Modeling

Now on the feature model, we run our XGBoost model on the 7 above-stated features and extract the best performing features using XGBoost's explain feature importance metrics. Overall, we get that User and User Personality are most impactful in predicting an output. This shows that who a user is is more impactful in predicting sarcasm rather than contextual features in the model.

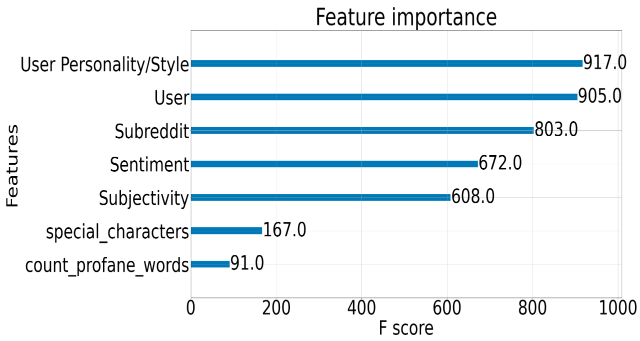


Figure 6: Feature Model Performance

Within the contextual features, we get subreddit and sentiment as the most important. This makes sense as different subreddits might be prone to different sarcasm just depending on the domain. Sentiment's impact makes sense as well as the tone of how someone talks can produce a feeling of "snark" and thus a sarcastic response.

In terms of the work distribution, Tusheet worked on the context-related features and analyzed features that have a higher correlation to sarcasm. He also performed statistical tests on these features to select the best ones for the feature-based prediction model. Mohan worked on training the BERT model and developing the BERT+CASCADE. He worked on training the feature-based model and performing the statistical tests and error modeling.

Conclusion

Overall, we have conducted an in-depth study over different models for predicting sarcasm using textual info. This includes using only the potential sarcastic response, only the context parent comment, or a combination of them both. We first show that the BERT model does well for a narrow and difficult field in NLP Classification in Sarcasm! We have provided a new baseline in predicting sarcasm by improving upon baseline and state-of-the-art models in sarcasm by achieving an 81% accuracy compared to the 77% performance by CASCADE. We do this utilizing a BERT+CASCADE model which takes advantage of BERT's wide NLP range and its power with Transformers. We have shown that more work needs to be done in combining both the parent and response comment in training for predicting sarcasm. The concatenation of parent and response comments proves to be noisy for predicting sarcasm. In this area, we propose utilizing two BERT encoders for parent and response separately to reduce the noise in future work.

Next, we do feature analysis and comparison. We analyze contextual features and find correlations, finding that special characters are correlated with sarcasm. We also extract user features and determine which features are most impactful in predicting sarcasm. We see that user-related features are a better indicator in predicting sarcasm compared to context-related features. Oddly enough, we see that special characters don't do a very good job of predicting sarcasm, showing that our initial high correlation didn't correspond with high importance.

For future work, we want to train on different datasets - especially those that have more explicit user features. We want to train models to generate longer contextual related features that match the user dimensional feature-length. This can allow us to utilize longer feature dimensions to train our XGBoost model. Lastly, we want to embed contextual and subreddit-related features into our BERT+CASCADE model in addition to user features as we see that contextual features such as sentiment and subjectivity still have a high importance score in predicting sarcasm.

Code Repository

Our code and analysis are on GitHub. Our dataset was too large to host on GitHub, so we have linked the dataset we used on Kaggle. Below are the links you need.
GitHub - https://github.gatech.edu/mdodda3/CSS_Sarcasm
Dataset - https://www.kaggle.com/datasets/mohandodda/wgca-user-embeddings

References

Amir, S., Wallace, B., Lyu, H., and Silva, P. 2016. Modeling 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 the Ninth International AAAI Conference on Web and Social Media.

Batista, J.M., Barros, L.S., Peixoto, F.V. and Botelho, D., 2022. Sarcastic or Assertive: How Should Brands Reply to Consumers' Uncivil Comments on Social Media in the Context of Brand Activism?. *Journal of Interactive Marketing*, *57*(1), pp.141-158.

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.

Dietterich, Thomas G. 1998. Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. Neural Computation., pages 1895-1923

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

Syafruddin, S., Thaba, A., Rahim, A.R., Munirah, M. and Syahruddin, S., 2021. Indonesian people's sarcasm culture: an ethnolinguistic research. *Linguistics and Culture Review*, *5*(1), pp.160-179.

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.