Context Based Sarcasm Detection

Project Report R3:

Team Members:

1. Arun Gaonkar (unity ID: agaonka)
2. Manasi Sanjay Ghosalkar (unity ID: mghosal)

1. Introduction

Sarcasm is a convoluted form of expression where meaning is conveyed implicitly. Recognizing sarcasm in a conversation is important for understanding the actual meaning and sentiment conveyed. The intended meaning is often different from what can be perceived by naive systems. This poses problems to many natural language systems, in particular, summarization, dialogue systems, and question-and-answering. It also has applications in understanding sentiments and opinions in modern communication channels such as tweets, comments, and chatbots.

Sarcasm detection remains a difficult task and this can be partly attributed to the fact that sarcasm relies heavily on the context of the dialogue taking place. Therefore, for this project, we compare the effect of having a reference context to detect the sarcastic intent of a piece of text as opposed to just the piece of text in consideration. We implement this approach on the News Headlines dataset which provides a set of sarcastic and non-sarcastic headlines followed by links to the news article. We use the text from these articles as context for the headlines.

2. Literature Survey

Sarcasm Detection has been explored by Davidov et al. (2010) [1] by applying semi-supervised techniques like SASI (Semi-supervised Sarcasm Identification Algorithm) along with feature extraction on two different datasets, consisting of tweets and product reviews. Gonzalez-Ibanez et al. (2011) [2] have also approached this task using supervised machine learning methods such as SVM and Logistic Regression. To address this problem in social-media domains, there have been works that deal with sarcasm in multi-modal settings such as texts and images (Schifanella et al. (2016) [3] and Cai et al. (2019) [4]).

Recurrent Neural Networks like LSTM (Long Short Term Memory) networks with sentence-level attention have been used by Ghosh et al. (2018) [5] by taking into consideration the sentences as well as the conversation context that the sentence responds to. This work shows that modeling the conversation context using a multiple-LSTM architecture yields better results in sarcasm detection as compared to just modeling the text in question. This argument is supported by other papers such as by Wallace et al. (2014) [6] which shows that even human annotators have to rely on additional context in order to classify or deduce ironic content. This has also been applied by others such as Pant et al. (2020) [7] using transformer model-based approaches.

3. Dataset

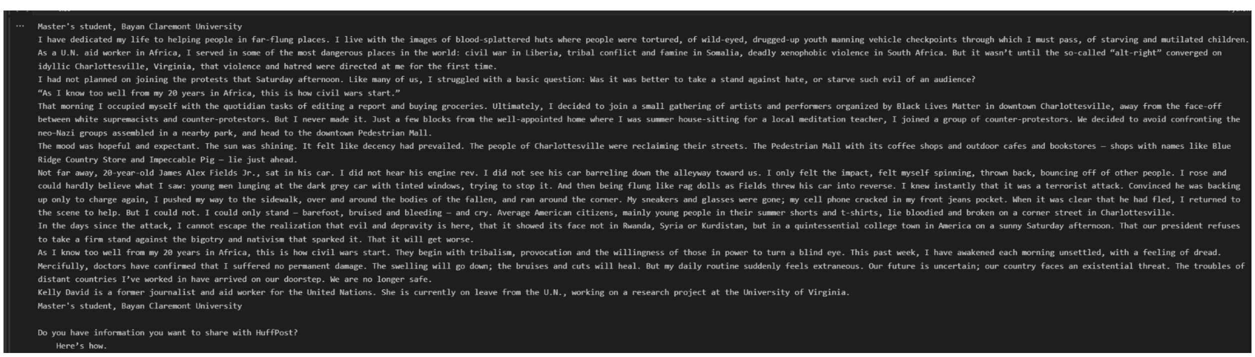
We have used the News Headlines for Sarcasm Detection dataset obtained from Kaggle [8,9]. The dataset is a json file that contains 28619 records. Each of these records are in the following format:

```
{
    “is_sarcastic”: 1,
    “headline” : “study: 83% of marathon spectators only attend for sick thrill of watching fellow man suffer”
    “article_link”:
"https://www.theonion.com/study-83-of-marathon-spectators-only-attend-for-sick-1828946111"
}
```

In each record,

1. The first item is the label. It is labeled 1 for sarcastic sentences and 0 for non-sarcastic sentences.
2. headline is the news article headline in the string format.
3. Last item is the article link, which hyperlinks to news articles in the form “https://…“. To get the context of the news, the article link can be used.

From the dataset description it is verified that the ‘headline’ from the article link and the headline from the record are matching. Most of the news articles are from sources like “TheOnion.com”, “huffingtonpost.com”. The Huffington Post is an American news website and political blog. The Onion is an American digital media company and newspaper organization that publishes satirical and sarcastic articles. For example, in the above example the article_link theOnion.com navigates to this website. And another example from huffpost.com navigates to
https://www.huffpost.com/entry/charlottesville-car-attack_b_5995ddd3e4b01f6e801ce11a
Each article has a brief description. We are using BeautifulSoup and html parser to extract the paragraphs in the article. And these paragraphs are used as context for the headlines.



```
Master's student, Bayan Claremont University
I have dedicated my life to helping people in far-flung places. I live with the images of blood-splattered huts where people were tortured, of wild-eyed, drugged-up youth manning vehicle checkpoints through which I must pass, of starving and mutilated children.
As a U.N. aid worker in Africa, I served in some of the most dangerous places in the world: civil war in Liberia, tribal conflict and famine in Somalia, deadly xenophobic violence in South Africa. But it wasn’t until the so-called “alt-right” converged on idyllic Charlottesville, Virginia, that violence and hatred were directed at me for the first time.
I had not planned on joining the protests that Saturday afternoon. Like many of us, I struggled with a basic question: Was it was better to take a stand against hate, or starve such evil of an audience?
“As I know too well from my 20 years in Africa, this is how civil wars start.”
That morning I occupied myself with the quotidian tasks of editing a report and buying groceries. Ultimately, I decided to join a small gathering of artists and performers organized by Black Lives Matter in downtown Charlottesville, away from the face-off between white supremacists and counter-protestors. But I never made it. Just a few blocks from the well-appointed home where I was summer house-sitting for a local meditation teacher, I joined a group of counter-protestors. We decided to avoid confronting the neo-Nazi groups assembled in a nearby park, and head to the downtown Pedestrian Mall.
The mood was hopeful and expectant. The sun was shining. It felt like decency had prevailed. The people of Charlottesville were reclaiming their streets. The Pedestrian Mall with its coffee shops and outdoor cafes and bookstores – shops with names like Blue Ridge Country Store and Impeccable Pig – lie just ahead.
Not far away, 20-year-old James Alex Fields Jr., sat in his car. I did not hear his engine rev. I did not see his car barreling down the alleyway toward us. I only felt the impact, felt myself spinning, thrown back, bouncing off of other people. I rose and could hardly believe what I saw: young men lunging at the dark grey car with tinted windows, trying to stop it. And then being flung like rag dolls as Fields threw his car into reverse. I knew instantly that it was a terrorist attack. Convinced he was backing up only to charge again, I pushed my way to the sidewalk, over and around the bodies of the fallen, and ran around the corner. My sneakers and glasses were gone; my cell phone cracked in my front jeans pocket. When it was clear that he had fled, I returned to the scene to help. But I could not. I could only stand – barefoot, bruised and bleeding – and cry. Average American citizens, mainly young people in their summer shorts and t-shirts, lie bloodied and broken on a corner street in Charlottesville.
In the days since the attack, I cannot escape the realization that evil and depravity is here, that it showed its face not in Rwanda, Syria or Kurdistan, but in a quintessential college town in America on a sunny Saturday afternoon. That our president refuses to take a firm stand against the bigotry and nativism that sparked it. That it will get worse.
As I know too well from my 20 years in Africa, this is how civil wars start. They begin with tribalism, provocation and the willingness of those in power to turn a blind eye. This past week, I have awakened each morning unsettled, with a feeling of dread.
Mercifully, doctors have confirmed that I suffered no permanent damage. The swelling will go down; the bruises and cuts will heal. But my daily routine suddenly feels extraneous. Our future is uncertain; our country faces an existential threat. The troubles of distant countries I’ve worked in have arrived on our doorstep. We are no longer safe.
Kelly David is a former journalist and aid worker for the United Nations. She is currently on leave from the U.N., working on a research project at the University of Virginia.
Master's student, Bayan Claremont University

Do you have information you want to share with HuffPost?
    Here’s how.
```

Figure 1: Sample Context extracted from the article_link

For example, the above image shows the parsed result of the hyperlink “https://www.huffpost.com/entry/charlottesville-car-attack_b_5995ddd3e4b01f6e801ce11a”.
When extracting article content, some paragraphs extracted were found to have only descriptions such as author details and dates. Such paragraphs with word length less than __ are discarded from the output as they do not hold any relevant article content. Another point taken into consideration is the length of an article as some pieces have more than 1000 words. However, due to the general structure of a news article, it can be said that the first few sentences would have enough information to summarize the content of the article. The articles were therefore truncated. The intuition behind this is that usually abstracts and overviews are mentioned in the first paragraph.

In the end, the dataset consisted of labels, article headline, and the article text. The task is to predict whether a headline is sarcastic or not with respect to a news article.

4. Methodology

For this project, we are comparing two different approaches. First approach is to use just the headlines from the dataset individually to predict whether they are sarcastic or not. This approach does not make use of any additional context and can be seen as a baseline approach. For the second approach, we are using both the headline and the article text to classify the headlines. The second method uses the headline followed by the article as input to the classification model. The inclusion of article text in the input adds additional context which is expected to provide better results than the baseline approach. We implement both these approaches on two models: LSTM using GLoVe [10] embeddings and the RoBERTa [11] classification model. Finally, we compare four resulting models.

For the first model, the sentences have to be converted into their vectorized format before feeding into LSTM, which is a type of Recurrent Neural Network (RNN). We use GLoVe embeddings called 'common crawl' which is 1.75GB in size and has 42B tokens to convert the sentences into word embedding inputs to the model. The model uses two Bidirectional LSTM layers followed by two Dense layers. For the Dense layers, the first layer uses a ReLu activation function while the final output layer has a sigmoid activation. To account for overfitting, Dropout layers are used.

For our second model, we have used a RoBERTa-base classification model from simpletransformers. RoBERTa is a transformer-based model previously trained on large-corpus of English data. This model was pre-trained with the objective of Masked Language Modeling (MLM) where 15\% of words in a sentence are randomly masked—and then running the entire masked sentence through the prediction model. Unlike recurrent neural networks (RNN) or GPT, the RoBERTa model learns the bidirectional representation of the sentence. This approach has been explored in Online Discourse in Dandu and Pant in [7]. The RoBERTa -base model has been trained for 5 epochs on GPU, keeping the maximum sequence length to be 256. The trained model was further evaluated on the test dataset for accuracy, F-1 score, and other metrics.

5. Evaluation

The prediction results of all the above-mentioned approaches on the test set are summarized in Table 1. With the Glove embedding and LSTM approach, 81% accuracy is achieved considering only the headlines. Combining the headline and the article text, we can see an increase in the performance by 15% in all metrics.
Transformers have better prediction results than the LSTM approach. This can be explained based on how transformers are handling long-term dependencies or because of their attention heads. The RoBERTa model, considering only the headlines, is 94% accurate. But including the article text separated by a separator token, an increase of 5.94% in F-1 score can be observed. We observe a similar significant increase in other metrics such as Precision, Recall, and Accuracy.

Model	Input format	Precision	Recall	F-1 Score	Accuracy
LSTM	**Headline**	0.8134	0.8167	0.8078	0.8079
LSTM	**Headline + article_text**	0.9592	0.9631	0.9607	0.9611
RoBERTa	**Headline**	0.9438	0.9359	0.9389	0.9402
RoBERTa	**Headline [SEP] article_text**	0.9984	0.9983	0.9983	0.9983

Table 1: Experimental Results on test set

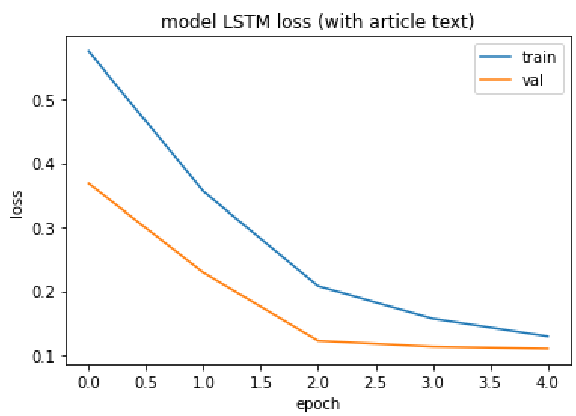


Figure 2.a

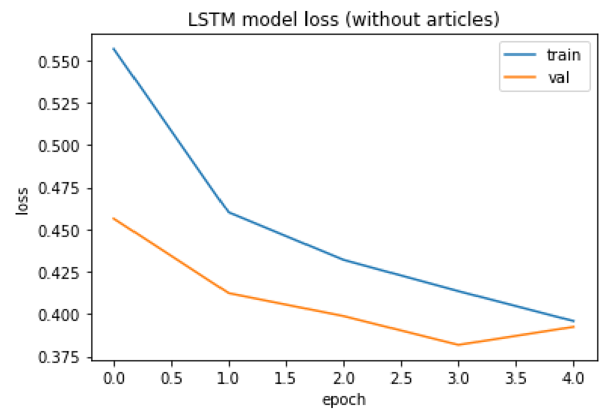


Figure 2.b

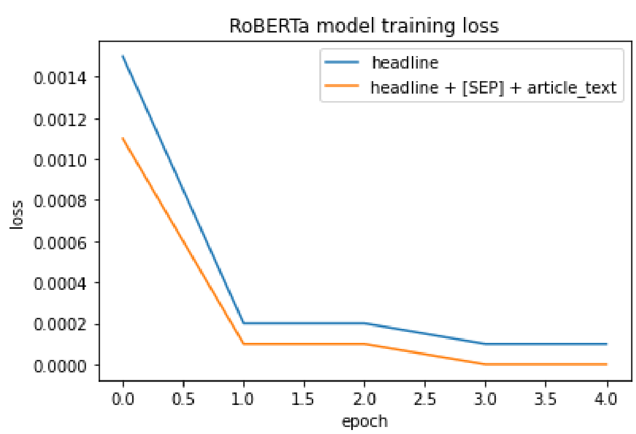


Figure 2.c

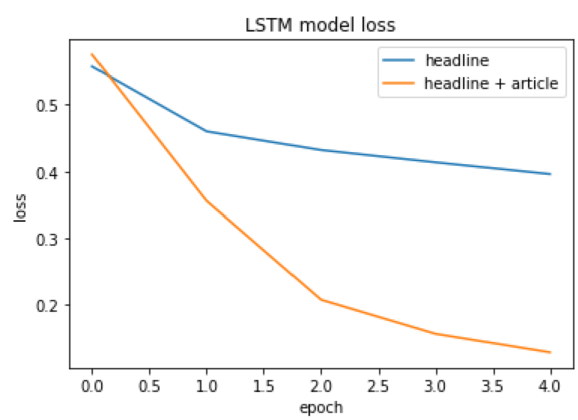


Figure 2.d

We investigated the training loss for all three approaches. From the loss curve, the loss stabilizes after training for 5 epochs in both LSTM and RoBERTa models. From figures 2 (a), (b) and (c), we observe that the initial training loss for RoBERTa is much lower than LSTM can achieve even after training for 5 epochs. Figure 2.(c) shows the comparison of training loss between the inputs being only headlines and headline [SEP] arcticle_text. Similarly the Figure 2.(d) shows the comparison of training loss between the different inputs given for the LSTM model.

6. Conclusion

Across both models, LSTM and RoBERTa, models which included article text along with the headlines were found to work better. This can be attributed to the fact that the article text provides additional context and information that acts as a basis to determine whether the given headline can be classified as sarcastic. This essentially solidifies the idea that context can prove useful in determining the sarcastic intent behind a piece of text.

Among the two proposed models, RoBERTa performed better than LSTM. This can be explained because transformers can handle long-range dependencies better than LSTM. Along with that, transformers are capable of capturing contextual information with minimal informational loss.

References

1. Dmitry Davidov, Oren Tsur, and Ari Rappoport. 2010. Semi-supervised recognition of sarcasm in twitter and Amazon. In Proceedings of the Fourteenth Conference on Computational Natural Language Learning, pages 107–116, Uppsala, Sweden. Association for Computational Linguistics https://aclanthology.org/W10-2914/
2. Roberto Gonzalez-Ibanez, Smaranda Muresan, and Nina Wacholder. 2011. Identifying sarcasm in twitter: A closer look. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, pages 581–586, Portland, Oregon, USA. Association for Computational Linguistics. https://aclanthology.org/P11-2102/
3. Rossano Schifanella, Paloma Juan, Joel Tetreault, and Liangliang Cao. 2016. Detecting sarcasm in multimodal social platforms.(https://dl.acm.org/doi/10.1145/2964284.2964321)
4. Yitao Cai, Huiyu Cai, and Xiaojun Wan. 2019. Multimodal sarcasm detection in twitter with hierarchical fusion model. pages 2506–2515. https://aclanthology.org/P19-1239/
5. Debanjan Ghosh, Alexander Fabbri, and Smaranda Muresan. 2018. Sarcasm analysis using conversation context. Computational Linguistics, 44:1–56. https://direct.mit.edu/coli/article/44/4/755/1620/Sarcasm-Analysis-Using-Conversation-Context
6. 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 Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pages 512–516, Baltimore, Maryland. Association for Computational Linguistics.
7. Tanvi Dadu and Kartikey Pant. 2020. Sarcasm Detection using Context Separators in Online Discourse. In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 51–55, Online. Association for Computational Linguistics. https://aclanthology.org/2020.figlang-1.6.pdf
8. Misra, Rishabh and Prahal Arora. "Sarcasm Detection using Hybrid Neural Network." arXiv preprint arXiv:1908.07414 (2019).
9. Misra, Rishabh and Jigyasa Grover. "Sculpting Data for ML: The first act of Machine Learning." ISBN 9798585463570 (2021).
10. Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. GloVe: Global Vectors for Word Representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1532–1543, Doha, Qatar. Association for Computational Linguistics.
11. Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, Veselin Stoyanov. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach https://arxiv.org/abs/1907.11692