Context Based Sarcasm Detection on News Headline Dataset

Arun Gaonkar
Department of Computer Science
North Carolina State University
Raleigh, NC, USA
agaonka@ncsu.edu

Manasi Sanjay Ghosalkar
Department of Computer Science
North Carolina State University
Raleigh, NC, USA
mghosal@ncsu.edu

***Abstract*—Sarcasm detection is an important aspect of natural language understanding in a social context. A sentence can essentially be interpreted in an entirely opposite manner than intended if not looked at in reference to the prior conversation or existing knowledge. This makes it highly contextual. Sarcasm detection is closely related to sentiment analysis in the sense that the sentiment of text depends on the context and situation, similar to sarcasm. So it is important to detect sarcasm in texts to understand the exact meaning the person is conveying.**

In this project, we attempt to evaluate the effect of prior context for detecting sarcasm in text sequences. For this, we are using the News Headlines dataset [1] [2] which has headlines labeled sarcastic and non-sarcastic along with the URLs to the corresponding news articles. The content of the article will be extracted and used as context for the headlines. It will then be used to predict the sarcasm in the headline using a pre-trained RoBERTa classifier. In this paper, we compare the prediction results of using only headlines with that of results when headlines are combined with the context by a separator token.

***Index Terms*—sarcasm, tokenization, GloVe, Bi-LSTM, RoBERTa**

I. 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.

II. Previous Work

Sarcasm Detection has been explored by Davidov et al. (2010) [3] 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) [4] 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) [5] and Cai et al. (2019) [6]).

Recurrent Neural Networks like LSTM (Long Short Term Memory) networks with sentence-level attention have been used by Ghosh et al. (2018) [7] 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) [8] 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) [9] using transformer model-based approaches.

III. 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 embeddings [10] and the RoBERTa [11] classification model. Finally, we compare four resulting models.

IV. Implementation

A. Data and Preprocessing

We have used the News Headlines for Sarcasm Detection dataset [1] [2] obtained from Kaggle. The dataset is a json file consisting of news headlines along with URLs to the news articles and a label 'is_sarcastic' indicating if the headline is sarcastic or not. Following is a sample data point:

{
"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"
}

For our approach to include additional context as input to the model, we have extracted the article text from the webpages the URLs pointed to using BeautifulSoup. Data cleaning steps were applied while extracting the text to discard irrelevant content like author details and date of publishing.

B. Model

- 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 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 [9]. 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.

C. Results

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.

TABLE I
Experimental Results on Test Set

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

V. 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] Misra, Rishabh and Prahal Arora, "Sarcasm Detection using Hybrid Neural Network." arXiv preprint arXiv:1908.07414 (2019).

[2] Misra, Rishabh and Jigyasa Grover. "Sculpting Data for ML: The first act of Machine Learning." ISBN 9798585463570 (2021).

[3] 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/

[4] 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/

[5] 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

[6] 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/

[7] 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

[8] 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.

[9] 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.

[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