Unmasking Sarcasm - Enhancing Sentiment Analysis in E-Commerce Reviews and Questions

Hugo Bouy
Illinois Tech
hbouy@hawk.iit.edu

Rémi Kalbe
Illinois Tech
rkalbe@hawk.iit.edu

Mathias Roumane
Illinois Tech
mroumane@hawk.iit.edu

1. Problem description

In an increasingly e-commerce driven world, the analysis of review, comments and questions written by consumers online has become a field of interests. Understanding opinions and emotions expressed in product feedback/related questions is essential to enhance the user's experience. One aspect that might be overlooked in this area is sarcastic and humor detection that can lead to a misinterpretation of these texts.

Humor remains a complex human phenomenon that is far from having a clear definition. While humor and sarcastic mechanisms are integral to human interaction, its subjective nature makes it a challenging target for computational analysis. Recent work has been able to open up this area using deep learning and natural language processing advances. With this project, we will attempt to improve e-commerce review processing using deep learning models for humor disambiguation.

2. Datasets

2.1. Combined dataset for sarcasm detection

To train this model, we used a collection of dataset of sentences. The main datasets used in the scope of this project is given below.

- **Headlines dataset**: Contains a list of 28,619 headlines collected from two news website. On one hand, TheOnions aims at producing sarcastic versions of real news events. On the other hand, real and non sarcastic news headloines are collected form HuffPost. This dataset has the advantage of having no spelling mistakes and informal usage since it is written by dedicated professionals in a formal manner.

- **MUStARD++ dataset**: Mustard++ is a multimodal sarcasm dataset that has been anotated whith 9 emotions. It was compiled form popular TV shows such as Friends, The Golden Girls or The Big Bang Theory. We will be using this dataset mostly to detect sarcasm but if we have time we may use the annotation to classify the emotion associated with the sarcastic sentence considered.

- **Sarcasm Corpus V2 dataset**: This dataset contains both sarcastic and non-sarcastic utterances. They are additionally classified in three different types: generic (6,520 samples), hyperbole (1.164 samples) and rhetorical (1,702 samples).

- **Sarcasm Amazon reviews dataset**: Contains a large number of both regular and ironic Amazon reviews. Each review is also associated with informations about the product for which the review was written, the number of stars assigned by the author, etc. This dataset will be the most usefull in the secind part of our project when we link the sarcastic review with the information regarding the product.

As an initial data processing step, we first concatenated those datasets together in order to use them to implement an initial sarcasm detection model. The final combined dataset then has the full sentence or review and a sarcastic indicator (0 or 1). The final combined dataset contains 40,461 samples from different sources and types. We then have a quite complex dataset with sentences of very variable length as displayed in figure 1.

2.2. Amazon reviews dataset

3. Results

3.1. Simple Sarcasm detection model

As a starting point for this project, we first aimed at creating a simple model capable of detecting if a sentence contains some form of sarcasm. At this stage, the model only takes into account the sentence and no additional factors. We were able to later on add additional relevant factors in detecting sarcastic or humorous patterns such as information on the product considered.

Our best results were obtained using the BERT pretrained model. BERT or "Bidirectional Encoder Represen-

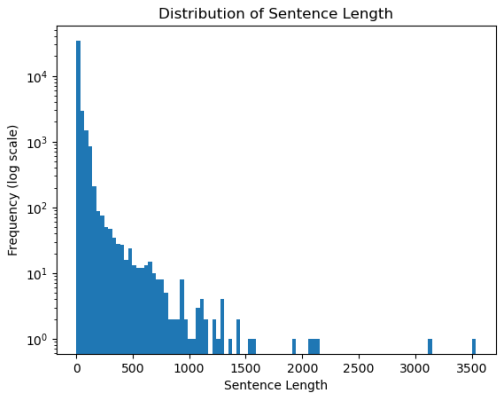

Figure 1. Frequency of sentence length in the final combined dataset

tations from Transformers", is a widely used pretrained natural language processing model developed in 2018. This model is built on a transformer architecture which enables efficient handling of sequential data. Additionally, it has already been trained on a massive amout of text data making an ideal choice for any specific NLP task after fine-tuning. Additionally, BERT uses a bidirectional approach making it possible to consider the entire context of a word in any given sentence making it better at capturing the nuance and complexity of laguage. Those few characteristics makes this state-of-the-art deep neural network ideal to capture the intricacies of a complex human behaviour such as sarcasm.

3.2. Sarcasm detection with additional factors

3.3. Comparison with Chat-GPT

4. What is left to be done

Our project focus on how sentiment analysis tools dedicated to sarcasm and humor disambiguation can be used to enhance E-commerce review and Q&A processing. It can be divided in 5 main milestones:

1. First, we implement a basic model able to detect humor/sarcasm in small text reviews. This model would only be capable of saying whether or not a given phrase/text contains humor or sarcasm patterns. To build the model, we compare different approaches published in the scientific literature.

2. Second, we improve the model to take into account several factors such as the type of product considered, its description, price, etc. (as sarcasm may be related to one aspect of the product). Feeding the model with the context can also help identify if some type of products are more prone to trigger humorous or sarcastic comments.

3. Third, we compare our model to a LLM such as Chat-GPT to assess if our dedicated humor/sarcasm classification model performs significantly better than a state of the art LLM using an appropriate prompt. We analyze the cost of both solutions.

4. Fourth, we improve our model to properly classify a product review considering its sarcastic/humorous sense into meaningful categories such as "consumer thinks the product is overpriced", "consumer thinks the product is good quality", etc. For instance, humor regarding the price of a product may indicate the product is overpriced compared to the customer's expectations.

5. If time allows it, we will use our sarcastic model detection to generate relevent automatic answer that take into account the humorous tone of the comment/question.

References

[1] Issa Annamoradnejad and Gohar Zoghi. Colbert: Using bert sentence embedding in parallel neural networks for computational humor, 2020. arXiv:2004.12765, updated Dec 2022.

[2] Aishwarya Gupta, Avik Pal, Bholeshwar Khurana, Lakshay Tyagi, and Ashutosh Modi. Humor@iitk at semeval-2021 task 7: Large language models for quantifying humor and offensiveness. arXiv:2104.00933, accepted at SemEval 2021.

[3] Sahil Jain, Ashish Ranjan, and Dipali Baviskar. Sarcasm detection in amazon product reviews. *IJCSIT*, 2018.

[4] Soujanya Poria, Erik Cambria, Devamanyu Hazarika, and Prateek Vij. A deeper look into sarcastic tweets using deep convolutional neural networks. *COLING*, 2016. arXiv:1610.08815.

[5] Hamed Yaghoobian, Hamid R. Arabnia, and Khaled Rasheed. Sarcasm detection: A comparative study, 2021. arXiv:2107.02276.

[6] Yftah Ziser, Elad Kravi, and David Carmel. Humor detection in product question answering systems. *Amazon Science*, 2020. Conference: SIGIR 2020.