Sarcasm Analysis - Difficulties & Unintended Benefits

Sentiment analysis refers to the analysis of natural language either in text or audio to extract, identify, and quantify affective states and subjective information. Sentiment analysis is a form of opinion mining that is most commonly applied to text based mediums such as product reviews or social media posts. Sentiment analysis, and derivative works, are also seen in domains such as financial markets or even in terrorism preventation [2]. Sarcasm analysis is a niche derivative application that attempts to detect sarcasm being used in text segments. There is one obvious benefit in detecting sarcasm which facilitates improving sentiment evaluation by disambiguating sarcastically “positive” text segments from truly positive text segments. Notably, there is another benefit that might unfold from such a technology that we have experienced in other contexts. Accessibility related technologies, such as closed-captioning, elevators, and ramps, have had fantastic unintended benefits for society outside of peoples with disabilities. This review will focus less on technological advancements but rather on some difficulties surrounding sentiment analysis and sarcasm detection as well as some more historical benefits that might be found tangentially from accessibility based technologies.

Background

Sentiment analysis has become a hot topic in the realm of natural language processing. As modalities of information are explored, typically the most obvious and easily extracted versions are leveraged first. Consider even the evolution of text retrieval systems where N-gram models and simple statistical/probabilities techniques are widely considered to be effective for most contexts. Sentiment analysis is readily being used as a means of augmenting traditional information modalities in industries such as marketing, finance, and national security.

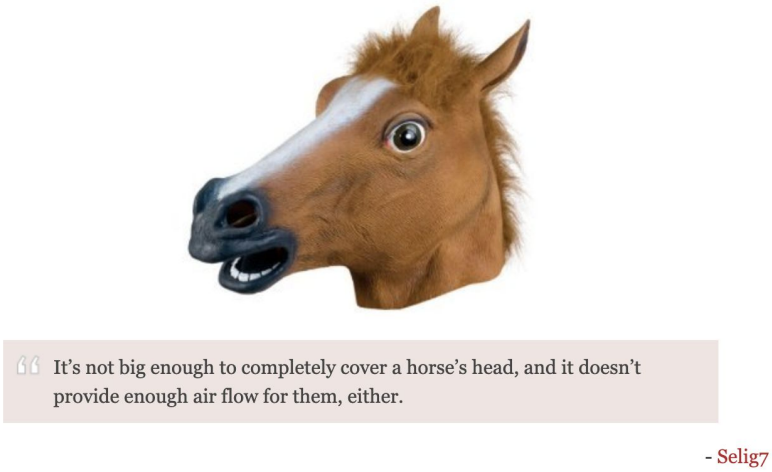
Sentiment refers to “what one feels about something”, “an attitude toward something” or “an opinion”. Sentiment analysis is inherently a measure of subjective qualities as one’s affective states are not solely expressed via quantitative means. The generalized process for sentiment analysis can be seen to the right. Effectively it comes down to three main steps: 1) parsing and cleaning the data, 2) transforming the text segments into more informative embedding models, 3) classifying the text segments.

Sourced from [1]

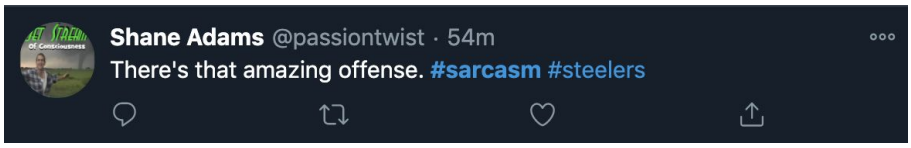
Sarcasm analysis can be viewed as a derivative of sentiment analysis as it relies on the same fundamental concepts - evaluating a subjective concept by using quantifiable and numerical methods. Sarcasm is a subtype of sentiment where one is trying to evaluate the expression of someone's feelings, either positive or negative, when the facade is that of the opposite feeling. Sarcasm is difficult to detect in text based mediums compared to conversational contexts where "tonal stress and certain gestural clues" can indicate sarcasm. In text based data, these clues are missing and thus ambiguity of meaning is increased.

Difficulties of Sarcasm Analysis

Sarcasm analysis is riddled with its own domain caveats, each posing a different problem to overcome. Consider the below examples of sarcasm in a few contexts: Amazon product views and a Twitter post.



Sourced from [4]



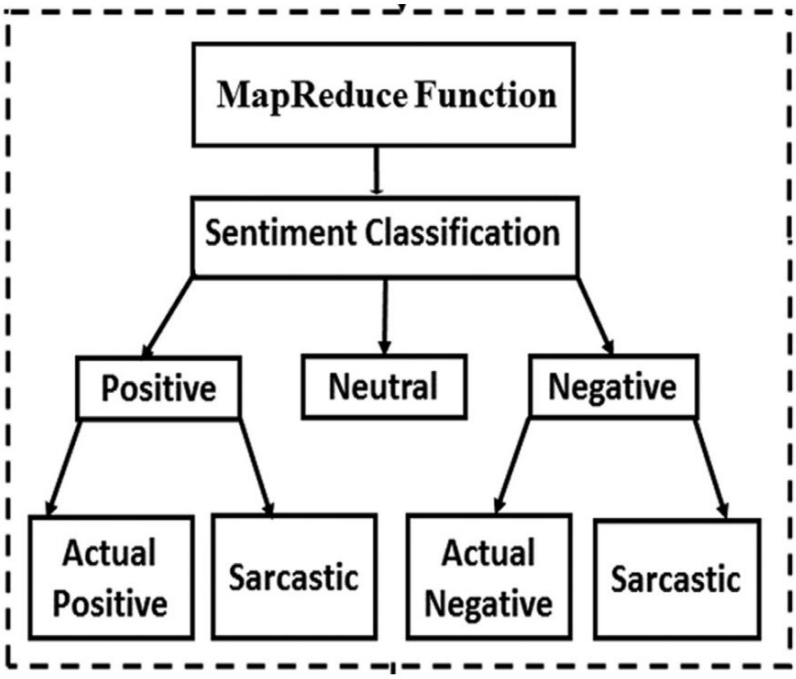
Sourced from [5]

Both examples highlight a few subtleties relating to the problem of sentiment analysis in natural language mediums. In the Amazon product review example, one might note that a simple context free analysis might lean towards this review being negative. However, we can understand that given the proper context that this is actually a "neutral" review, where it is neither positive nor negative - it is unrelated to some extent. In the twitter example, we have a case where a positive connotation would be detected given the usage of "amazing", however, a negative sentiment is actually being portrayed. The juxtaposition of these two examples highlights a few important concepts as it relates to both sarcasm and sentiment analysis. One, not every text segment will be marked as sarcastic, either via a hashtag, an emoji, or some

other hyperbolic portions; two, there are many dimensions to the classification of a given segment: sarcasm on a negative comment is not always positive and vice versa.

While not covering every modality of sarcasm detection, the figure below highlights an approach to solve a caveat of sentiment analysis as it relates to sarcasm - trying to distinguish between true and false positive/negative sentiment classifications.

The main methodologies of sarcasm detection features a collection of both heuristic and strictly lexical approaches, as shown in the below figure. As shown in the above examples, one simple way of detecting sarcasm is utilizing text segments that are commonly used to indicate sarcasm such as “#sarcasm”, “/s”, or various emoticons. Sarcasm can also be indicated via usage of hyperbolic emphasis such as quotes or excessive punctuation. However, as shown in the Amazon product example, these obvious indicators are not always present which is when we need to rely on pure lexical means of analysis.

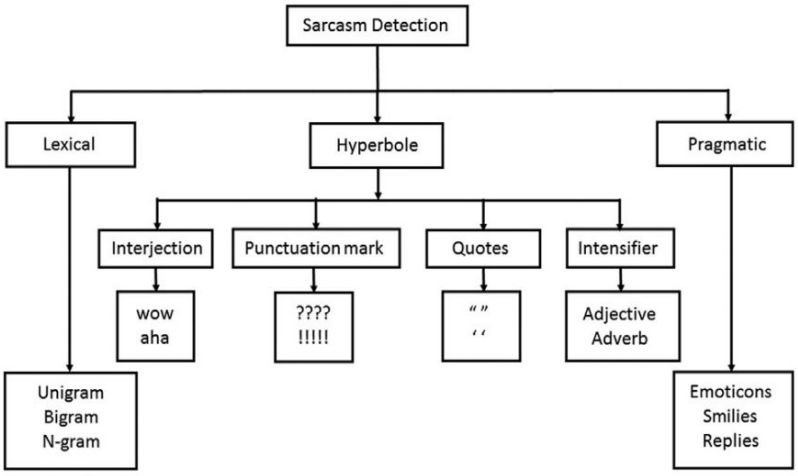


Sourced from [9]

Overview of Solution Spaces

Just like sentiment analysis and notably information retrieval systems, the processing steps in sarcasm detection are similar if not identical. Diving a bit deeper into the final classification step in sarcasm detection one can find a variety of solutions, each slightly tailored to their application space. Traditional machine learning approaches have been used to perform this classification, such as SVMs, Naive Bayes, Logistic Regression, and decision trees/forests [3].

Another set of solution spaces has been explored in the Deep Learning community utilizing architectures such as Convolutional Neural Networks and Recurrent Neural Networks [6]. More fully end-to-end approaches have been explored which utilize Deep Learning based word embeddings combined with Deep Long-Short Term Memory based neural networks [7].



Sourced from [9]

As mentioned, there are always unintended consequences and benefits to innovations and products. Accessibility innovations such as closed-captioning, elevators, and ramps were originally designed for people with disabilities. Consider today how often such services are used on a daily basis by individuals without disabilities. Closed-captioning is prominently used in airports, bars, and other noise-sensitive areas such as waiting rooms.The Americans with Disabilities Act (ADA) has provided invaluable quality-of-life improvements to those with and without disabilities since its inception [8].

Conclusion

WIth the proliferation of sentiment analysis applications in various technology spaces, one must be careful of the caveats associated - such as “is this sentiment actually genuine or not.” Sarcasm detection is a means to disambiguate such a concern and is extremely important in more informal text based mediums such as social media posts, customer reviews, or forum-based comments. The process of determining sarcasm has its own subtle caveats such as detecting sarcasm when no obvious lexical clues are provided. While not immediately obvious, sarcasm detection might evolve into an accessibility feature like closed-captioning that has incredible generalized benefits to society.

Citations

[1] “(Tutorial) Simplifying Sentiment Analysis in Python.” *DataCamp Community*, www.datacamp.com/community/tutorials/simplifying-sentiment-analysis-python.

[2] Ahmad, S., Asghar, M.Z., Alotaibi, F.M. et al. Detection and classification of social media-based extremist affiliations using sentiment analysis techniques. Hum. Cent. Comput. Inf. Sci. 9, 24 (2019). https://doi.org/10.1186/s13673-019-0185-6

[3] Singh, Digvijay. “Sarcasm Detection: Step towards Sentiment Analysis.” *Medium*, Towards Data Science, 10 July 2019, towardsdatascience.com/sarcasm-detection-step-towards-sentiment-analysis-84cb013bb6db.

[4] Cosgrave, Luke, and Le Conte Stewart. “A Review.” *Amazon*, Red Butte Press, 1997, www.amazon.com/review/R6PAMGK0Y6WMT/ref=azfs_379213722_AccoutrementsHorseHeadMask_4?pf_rd_m=ATVPDKIKX0DER.

[5] Adams, Shane. “There's That Amazing Offense. #Sarcasm #Steelers.” *Twitter*, Twitter, 15 Nov. 2020, twitter.com/passiontwist/status/1328117541559001090?s=20.

[6] S. Porwal, G. Ostwal, A. Phadtare, M. Pandey and M. V. Marathe, "Sarcasm Detection Using Recurrent Neural Network," 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2018, pp. 746-748, doi: 10.1109/ICCONS.2018.8663147.

[7] S. S. Salim, A. Nidhi Ghanshyam, D. M. Ashok, D. Burhanuddin Mazahir and B. S. Thakare, "Deep LSTM-RNN with Word Embedding for Sarcasm Detection on Twitter," 2020 International Conference for Emerging Technology (INCET), Belgaum, India, 2020, pp. 1-4, doi: 10.1109/INCET49848.2020.9154162.

[8] Shapiro, Joseph. “In Helping Those With Disabilities, ADA Improves Access For All.” *NPR*, NPR, 24 July 2015, www.npr.org/2015/07/24/423230927/-a-gift-to-the-non-disabled-at-25-the-ada-improves-access-for-all.

[9] Bharti, S.K., et al. “Sarcastic Sentiment Detection in Tweets Streamed in Real Time: a Big Data Approach.” *Digital Communications and Networks*, Elsevier, 12 July 2016, www.sciencedirect.com/science/article/pii/S235286481630027X.