Understanding Sarcasm and its Application in Artificial Intelligence through Detection in Plain Text

Jahn Carlo Marrero Santiago
Human Perception of Artificial Intelligence
Spring Semester 2018-2019
Prof. José Meléndez

Project Repository

To access project repository with all the needed files and instructions to run the method in this research project, go to: https://github.com/JahnCarlo/hpaiproject . A version of this document is also present in that repository.

Abstract

Sarcasm is a powerful tool when used right. Whether for comedic purposes or to convey that you mean the opposite of what you are saying, sarcasm is used in voice and text. However, it is sometimes really hard to detect sarcasm, because it varies per person. This means that two persons can have a different perception of the use of sarcasm, and then, when someone decides to use sarcasm in their speech, it can go undetected. This research project is meant to see what sarcasm is, how can it be detected and the different aspects of both spoken and written english language that define it. It is also meant to explore one of the many ways that it is applied to Artificial Intelligence, and how can Artificial Intelligence detect something that even humans do not detect sometimes. Finally, an attempt at a neural network will be made, and the results will be discussed.

Introduction

Humans use speech, among other things, to express their emotions or what they concieve to be their emotions. One of the emotions that is hard to sometimes communicate is sarcasm. As defined by Merriam-Webster, sarcasm is “a sharp and often satirical or ironic utterance designed to cut or give pain” or “a mode of satirical wit depending for its effect on bitter, caustic, and often ironic language that is usually directed against and individual”. This research project will be based on the latter, meaning it is focused to see what really is sarcasm, and then use some methods already implemented in the Python language and using Tensorflow by a group of students from Rochester Institute of Technology to see how can it be applied to Artificial Intelligence.

Research

Sarcasm as a Language Element

These days, sarcasm is an essential spoken element in our daily lives. The society we live in is heavily inclined towards the use of ironic messages to convey a certain emotion they are feeling and sarcasm is the perfect tool for that. Through sarcasm, we can express a message and be ironic at the same time. That is of course if the other person gets you are being sarcastic. For example, a parent may say to their child “This room is just as clean as I wanted you to have it by this morning” when the room is still dirty, and using a tone of mockery. The child then perceives the sarcasm and gets that the parent is mad at them for not doing their chore. What the parent wanted to say was that he/she is not content about the fact that the room is still dirty, but used totally opposite words to express it.

The difficulty of detecting sarcasm as a language element then depends on how each person uses sarcasm. It is argued by an article published by Richard Chin in the Smithsonian Magazine that "The mental gymnastics needed to perceive sarcasm includes developing a "theory of mind" to see beyond the literal meaning of the words and understand that the speaker may be thinking of something entirely different. A theory of mind allows you to realize that when your brother says "nice job" when you spill the milk, he means just the opposite, the jerk." This means that while you are listening to the sarcastic voice, in your head you start connecting the dots in why the message is being positive if the action was negative, or vice versa. You can maybe understand the flow of thought of sarcasm detection by going step by step on what you really think when you hear something sarcastic. If you've never heard a certain sarcastic expression like "Wow amazing job" when you spilled the milk, you then think "why would this be an amazing job if I made a mess in the kitchen", and then you come into retrospective with yourself and think that when the voice said "Wow amazing job", they really meant the total opposite. And for me this is exactly where the magic of sarcasm comes. You don't expect to be praised for doing wrong and people don't expect to hear someone had a pleasant night when they really got robbed that night. It's in the retrospective you have with yourself for a brief unit of time that you come to think of that sarcastic message that was conveyed to you, and you think of it differently.

Listening to sarcasm sparks an emotion in each of us differently. For example, let's take into account that spill event. The way you managed the situation would have been completely different if you were told "Oh no! You spilled the milk, I'll help you clean it". With this statement, you would probably feel guilt, but you would feel much better throughout the whole situation because no negative meaning was instilled in the sentence and therefore, you see the even as something insignificant. However, since sarcasm was used instead, you may then feel a lot of guilt, and next time you try to get milk when the person that used sarcasm is around, you will be more careful in spite of not wanting to hear that negative tone in sarcasm again.

The effect of sarcasm is directly linked to how humans perceive emotion, and how the brain uses predictions as a form of reflex. Dr. Lisa Feldman Barrett addresses predictions of the brain in her book *The Secret Life of The Brain* by stating that "Predictions are such a fundamental activity of the human brain that some scientists consider it the brain's primary mode of operation. Predictions not only anticipate sensory input from outside the skull but explain it" (Page 59). This means that, in our case, when someone says something sarcastic or we read something written with sarcastic innuendo, our brain is deceived of that prediction the first time we hear or read that sarcastic message. Whether we expected help when we spilled the milk or we didn't expect an appraisal when we did something wrong, makes our brain try to make sense of that unexpected event because our brain failed to predict what would really happen; it did not expect someone to say or write a reaction totally opposite of what we expected it to be. This then, through what Dr. Feldman calls *prediction loop* (Predict -> Simulate -> Compare -> Resolve Errors -> Repeat) (Page 63), our brain creates sarcasm. We predicted a negative reaction, the reality we perceived was a positive reaction, we compare using the situation (spilling the milk for example), we detect the error that we didn't perceive and therefore we classify that totally opposite reaction as sarcasm.

We can then argue that the detection of sarcasm in language depends on how our brain feels emotions. Sarcasm then can be classified as a combined responses of both the actions that took place in the reality you perceived and the emotions you felt in that space and time slot of the event. This makes sarcasm a complex language part and emotion that humans use in communication.

Sarcasm and Artificial Intelligence

To make a useful traduction of sarcasm in a way that it can be processed by Artificial Intelligence, further analysis in sentence structure has to be done. Typically, modern Artificial Intelligence can detect sarcasm on plain text. A study done by students at the University of Utah has a good approach of the different polarities in a sarcastic sentence structure. They define a sarcastic sentence to have completely different sentiments on only one sentence. They classify two parts of the sentence: a positive sentiment and a negative activity/state. Some of the examples given are (take positive sentiments to be underlined and negative activities/states italicized): "Oh how I love being *ignored*", "Absolutely adore when my *bus is late*", "I'm so pleased at how *my mom woke me up* vacuuming this morning!" and "I love *fighting* with the one I love". We can take this approach to then classify how opposite are things in a sentence, which opens multiple opportunities to apply sarcasm detection to plain text using artificial intelligence.

Then, naturally, it may come to question what kind of technology could be used to apply Artificial Intelligence to these kind of datasets (plain text of sentences) so that we can detect the presence or absence of sarcasm. Instinctively, one may think of Machine Learning as the source technology to apply this kind of analysis, since Machine Learning has proven over the years that is very intuitive to make software 'learn' in a similar way that humans do. There are some technologies that can be used with this purposes: Support Vector Machines and Deep Neural Networks are two of the Machine Learning branches we can dive into to see how each can be used to detect sarcasm in plain text.

A project done by student Miruna Pislar for the University of Manchester explored the detection of sarcasm in tweets using Supervised Vector Machines as one of her methods. Tweets are a very useful source because not only are they numerous, they can be even monitored live. In the study, Miruna took different steps into analyzing a tweet that could help her into detecting sarcasm. Firstly, a classification of certain features of a tweet like punctuation marks, expressions like "Oh" and "Wow", user mentions ,hashtags, emoticons and adjectives like "amazing" and "incredible" was done. This category was named Pragmatic Features because they were not really a part of speech, but they were used to judge the actual message of the tweet. This went in hand with using common sentiment analysis to detect parts of the sentence and parts of speech that indicated neutral, positive or negative sentiments in the sentences. This then resulted in a polarity score that reported percentages of all three neutral, positive and negative sentiments for the whole tweet. What is interesting is that some of the sarcastic tweets used in the research, showed that sentiment incongruity was present. This sentiment incongruity can be explained when words with negative and positive connotation were placed one beside the other; an example can be "... delayed! amazing! ...". Here the word delayed obviously reported a negative sentiment, but the word amazing reported a positive sentiment. This can help explain in a more elemental level what sarcasm is. It helps to explain that the nature of sarcasm is often using very different polarity of verbs and adjectives in sentences to explain a common feeling; often of mockery.

In addition to the characteristics of the tweets already mentioned, they tried to detect topics in the sentences and how some of the polarity of words related to those topics. All this data was then fed to a Support Vector Machine to do the training. A Support Vector Machine is a branch of machine learning and supervised models that given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier. The results of this model for them showed that the model generated by the SVM could take a sentence and successfully detect it was sarcastic 85% of the time. This means that although very close, it is very hard to really define sarcasm in the Artificial Intelligence world.

Methods

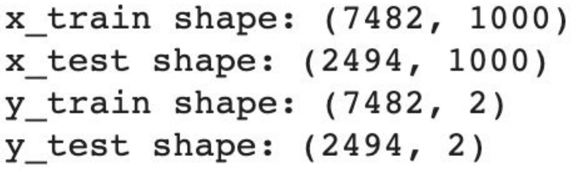
After I did the research compiled above, I wanted to implement a Natural Language Processing Neural Network that detected if a tweet was sarcastic or not. I used the Google Colab made by Emmanuel Dufourq titled Natural Language Procesing - Classification, which was made with the intention of classifying StackOverflow questions with a programming language. I used this colab as a base of mine, similar to a project that was made in the Human Perception of Artificial Intelligence class. The technologies used were Pandas, Keras, Numpy, NLTK and BS4. The use of the first three was already described in class, but NLTK and BS4 are both data processing libraries for python. NLTK helped to remove stop words from the data I was analyzing and BS4 helped to correct spelling in such tweets.

To make an efficient neural network, I needed a good dataset. For this, I found an existing website that uses a lot of Artificial Intelligence techniques to detect the percentage of sarcasm in a sentence. This website is called Sarcasm Detector, and it has an open source github page I could access. I went into his folder structure and found two twitter datasets of both sarcastic and regular, non-sarcastic tweets. These were two Comma Separated Values files, composed of more than fifteen thousand tweets each. I downloaded them and compiled eight thousand tweets from each. Then, I proceeded to tag the tweets with a 0 if they were not sarcastic and a 1 if they were, collecting a new Comma Separated Values file of sixteen thousand tweets.

As it may be known, the problem with using tweets as data set is that a tweet has a lot of elements that will be of no use such as mentions that start with '@', hashtags that start with '#' and other language. For this, I had to do extensive data processing, using regular expressions to filter out my text. After all this data processing, I had about five thousand sarcastic tweets and non sarcastic tweets, eliminating then the ones that were not useful.

After cleaning the tweets, the data needed to follow the original base code that Emmanuel wrote was almost done. I had to confirm I had only two distinct tags, 0 for non-sarcastic text and 1 for sarcastic text. This means that my Neural Network would have 2 classes to classify text, 0 or 1. Then, I used 25% of the tweet data set for testing the neural network, and the remaining 75% to train the neural network. I followed some tokenizing I had to do with a limit of one thousand words as it was in the original base code for Natural Language Processing.

Then, I inspected the shapes of each of my variables. These were the shapes of the variables:



```
x_train shape: (7482, 1000)
x_test shape: (2494, 1000)
y_train shape: (7482, 2)
y_test shape: (2494, 2)
```

This meant that a total of seven thousand four hundred and eighty two examples were used to train the neural network (75%), and that each consisted of a thousand words each. Also, the test data was two thousand four hundred and ninety two in size, and only had 2 dimensions : either sarcastic or not. At this point, I only needed the neural network.

The neural network used was very similar to the one Emmanuel had done, I just changed the epocs to 70 to allow more testing. The model used was sequential, which meant that it will run layer by layer, no convolution. The first dense layer had an input size of 512 and the shape was 1000 words which was defined before. Then it had a 'relu' (Rectified Linear Unit) activation. Relu is often used to avoid that the output of the dense is 0, so the gradient is not 0, so it can recover throughout the other layers. Then we added a dropout rate of 0.5 followed by a dense the size of the number of classes(2). This means that at this point, we can almost decide if the sentence is sarcastic or not. The next dense had a softmax activation layer, which means that it will output a number between 0 and 1 that will determine the probability of the sentence being sarcastic or non sarcastic. The model was compiled with categorical cross entropy loss (often used in natural language processing).

Results

After training and checking the accuracy of my model, I was amazed to see a 93.6% of accuracy. This meant that for 100 arbitrary tweets I chose, this model will predict presence or absence of sarcasm in 93 of them. This was great so I had to put it to the test. These are some sentences I typed in my model with my intent presence or absence of sarcasm and the predicted one.

Sentence	**Intent**	**Prediction**
I am so ready for this test NotReally	Sarcastic	Non Sarcastic
wow. I love going to the dentist early in the morning. great.	Sarcastic	Sarcastic
I love my dog	Non Sarcastic	Non Sarcastic
Im so happy the whole world crumbled on me. yikes.	Sarcastic	Non Sarcastic

Discussion

These were a small amount of sentences to try but I think other intents of putting actual sentences in the model would be futile. I believe the model gave 93% of accuracy, but it has to me remarked that this model was made to detect sarcasm in processed *tweets*. Tweets are not like sentences at all, they are short and not true to grammar. This means that one tweet can have multiple meanings; this is because grammar itself is very ambiguous. This model can successfully predict the sarcasm in a tweet, but not in a sentence, which brings me to think that there is more to detecting sarcasm in sentences or tweets than a simple sequential neural network for natural language processing, and at the same time brings sense to all the works I researched and all the technologies they used aside from neural networks to achieve an efficient detection. Not only analyzing and classifying which sentence could be sarcastic or not, but also analyzing parts of speech, words used, punctuation and polarity of sentiment of the sentences was always needed to make an efficient prediction.

Conclusion

Detecting human emotions using Artificial Intelligence is much harder than one can imagine. A lot of factors have to be taken into account: sentence sentiment polarity, punctuation marks; even context is somewhat very important. However, the closest we will every get to detecting the phenomenon of sarcasm in text will always be a prediction. We will always have a margin of being wrong in predicting sarcasm, and if Humanity wants to detect sarcasm in speech and mannerism, it becomes a problem of three dimensions, where one dimension adds difficulty to the previous one. I now appreciate the power of Machine Learning techniques such as Neural Networks, Supervised Vector Machines and others in the attempt and close achievement they bring to humanizing computer-human interaction, and making it more natural for automated services to interact with humans. As for sarcasm detection, further analysis of sentences will have to be done and datasets related to normal conversations will be more useful for everyday detection in normal conversations between AI and humans. I expect this technology to rocket over these few coming years and bring a drastic change to how computing and Artificial Intelligence is seen by humanity.

Acknowledgements

I want to take a space of this document to thank Dr. José L. Meléndez for bringing this class to UPRM and giving us a space to think about how complicated humans, and put us students in a space where we know how important it is to know how humans function first to then translate into the Artificial Intelligence world. One of the best things I take from this class is that not one human is like the next, and that many of our experiences, if not every single one of them, define who we are and how we think; this means that not one human is better than the other, which means that we have to work in unison to achieve a better understanding of reality and how it affects us.

References

Barrett, L. (2017). *How emotions are made - The Secret Life of The Brain*. 1st ed. Chapters 1 - 9.

Chin, R. (2011). *The Science of Sarcasm? Yeah, Right*. [online] Smithsonian. Available at: https://www.smithsonianmag.com/science-nature/the-science-of-sarcasm-yeah-right-25038/ [Accessed 10 Apr. 2019].

Cliche, M. (2013). *Sarcasm_detector*. [online] GitHub. Available at: https://github.com/MathieuCliche/Sarcasm_detector/tree/master/app [Accessed 10 Apr. 2019].

Khatwani, S. (2019). *Sarcasm_Detection_using_Tensorflow*. [online] GitHub. Available at: https://github.com/SanjayKhatwani/Sarcasm_Detection_using_Tensorflow/blob/master/sarcasm-detection-report.pdf [Accessed 10 Apr. 2019].

Merriam-webster.com. (2019). *Definition of SARCASM*. [online] Available at: https://www.merriam-webster.com/dictionary/sarcasm [Accessed 10 Apr. 2019].

Riloff, E., Qadir, A., Surve, P., De Silva, L., Gilbert, N. and Huang, R. (2019). *Sarcasm as Contrast between a Positive Sentiment and Negative Situation*. [online] Cs.utah.edu. Available at: https://www.cs.utah.edu/~riloff/pdfs/official-emnlp13-sarcasm.pdf [Accessed 10 Apr. 2019].