# “Ha, I *definitely* meant that!”: Detecting Sarcasm in Reddit Comments

Nanda H Krishna, Rubini U and Vikram Reddy

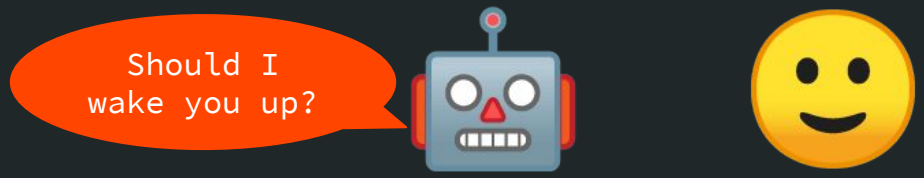
June 27, 2020

# Problem Statement

- Sarcasm – “the use of irony to mock or convey contempt”
- Widely employed in conversation
- Humans have a hard time detecting sarcasm
- Machines may find it even more difficult
- Important problem to solve for Natural Language Understanding
- Sarcasm probably required to pass Turing Test

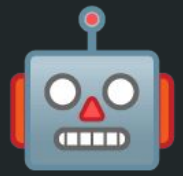
# Problem Statement

- Sarcasm – “the use of irony to mock or convey contempt”
- Widely employed in conversation
- Humans have a hard time detecting sarcasm
- Machines may find it even more difficult
- Important problem to solve for Natural Language Understanding
- Sarcasm probably required to pass Turing Test



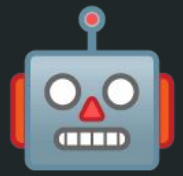
# Problem Statement

- Sarcasm – “the use of irony to mock or convey contempt”
- Widely employed in conversation
- Humans have a hard time detecting sarcasm
- Machines may find it even more difficult
- Important problem to solve for Natural Language Understanding
- Sarcasm probably required to pass Turing Test



# Problem Statement

- Sarcasm – “the use of irony to mock or convey contempt”
- Widely employed in conversation
- Humans have a hard time detecting sarcasm
- Machines may find it even more difficult
- Important problem to solve for Natural Language Understanding
- Sarcasm probably required to pass Turing Test

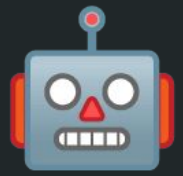


# Problem Statement

- Sarcasm – “the use of irony to mock or convey contempt”
- Widely employed in conversation
- Humans have a hard time detecting sarcasm
- Machines may find it even more difficult
- Important problem to solve for Natural Language Understanding
- Sarcasm probably required to pass Turing Test

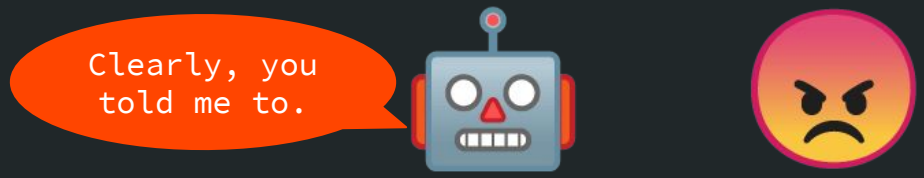

# Problem Statement

- Sarcasm – “the use of irony to mock or convey contempt”
- Widely employed in conversation
- Humans have a hard time detecting sarcasm
- Machines may find it even more difficult
- Important problem to solve for Natural Language Understanding
- Sarcasm probably required to pass Turing Test

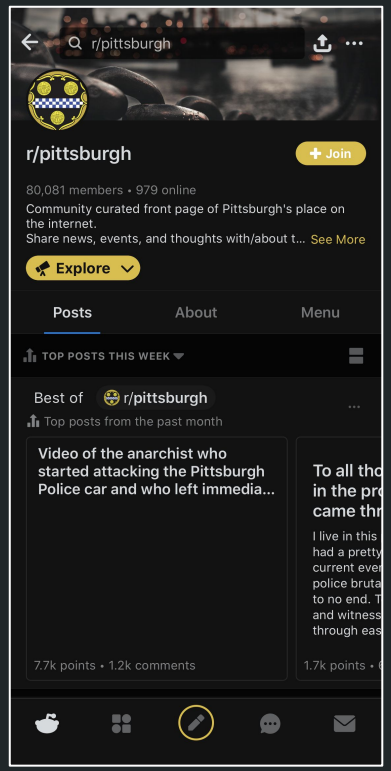


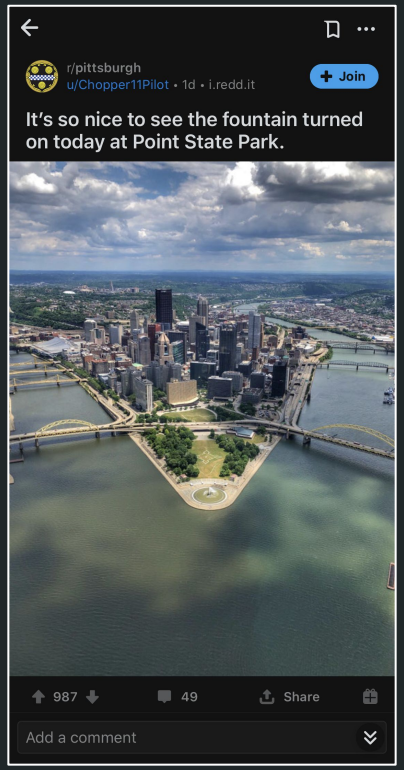
# Problem Statement

- Sarcasm – “the use of irony to mock or convey contempt”
- Widely employed in conversation
- Humans have a hard time detecting sarcasm
- Machines may find it even more difficult
- Important problem to solve for Natural Language Understanding
- Sarcasm probably required to pass Turing Test



# Reddit


r/pittsburgh
r/pittsburgh
Join
80,081 members • 979 online
Community curated front page of Pittsburgh's place on the internet.
Share news, events, and thoughts with/about t... See More
Explore
Posts
About
Menu
TOP POSTS THIS WEEK
Best of r/pittsburgh
Top posts from the past month
Video of the anarchist who started attacking the Pittsburgh Police car and who left immedia...
7.7k points • 1.2k comments


r/pittsburgh
u/Chopper11Pilot • 1d • i.redd.it
Join
It's so nice to see the fountain turned on today at Point State Park.
987
49
Share
Add a comment


r/pittsburgh
987
49
Share
gameboy_homeboy • 1d
God our city is beautiful
Reply
56
Tendie_Obtianer • 1d
No
-48
nickmcmillin • 1d
Elaborate
3
HugeFinish • 16h
Why?
Vote
cashewcheez • 1d
They turned on the fountain todayyy at Point State Park, and I felt something in me change
Reply
27
beard_tan • 1d
I was thinkin' about how we all feel, but the world's not such a shitty place.
7
Forsworn_Hon0r • 1d
We just can't blame the seasons; the Blue Man Group won't cure depression.
6

# Dataset

- Original dataset gathered by Mikhail Khodak et. al.
- 1.3 million sarcastic comments
- ‘/s’ in comments indicates sarcasm, but not always present
- Downsampled and balanced version available on Kaggle
- Further downsampled due to computational limitations
    - 125000 unique comments
    - Nearly equal number of sarcastic and non-sarcastic comments
    - 80-20 split for training and validation data

# Features

| | |
|---|---|
| Label | 0 |
| Comment | Hard to convey sarcasm over the internet |
| Author | dcpp4 |
| Subreddit | Amd |
| Score | 1 |
| Upvotes | 1 |
| Downvotes | 0 |
| Year and Month | 2016-09 |
| UTC Timestamp | 9/9/2016 00:31 |
| Parent Comment | You’re the only one |

# Methods Used

- Approach – extract embeddings for text data and pass it to a classifier
- Embeddings
  - TF-IDF (**T**erm **F**requency - **I**nverse **D**ocument **F**requency)
  - BERT (**B**idirectional **E**ncoder **R**epresentations from **T**ransformers)
- Classifiers
  - Ensemble models – Random Forest, Gradient Boosting
  - Gaussian Naïve Bayes
  - Neural Networks – `scikit-learn` Multi-Layer Perceptron, `Keras` ANN
- Dimensionality reduction using PCA

# Methods Used

- Strategies to handle text data
  - Pre-processed text (removal of punctuation and stopwords, lemmatisation)
  - Raw text
  - Using n-grams
- Features used
  - Comment
  - Parent
  - Statistics
- All 3 strategies used for TF-IDF, only raw text used for BERT

# Methods Used

- Hyperparameters
  - `max_features=1000 (or) 5000` for TF-IDF
  - `n_estimators=50` for Random Forest Classifier
  - `hidden_size=(25,)` for Multi-Layer Perceptron
  - ANN
    - 4 hidden layers (512, 256, 128, 64) and Fully-Connected Layer
    - Adam optimiser, binary cross-entropy loss
- The pre-trained BERT-Base model was used for the BERT Embeddings

# Evaluation

- Plain accuracy isn't enough
  - Precision and recall (class-wise) are important
  - We prioritise $F_1$-score (harmonic mean of precision and recall)
- Explaining model outputs
  - Need to identify patterns and features that indicate sarcasm
  - Seeing how the models understand language is important
  - Qualitative evaluation
  - Look at weights, activations, feature importances, etc.
  - Important to ensure fairness and robustness of the system

# Results

- TF-IDF features extracted from pre-processed text
  - Random Forest Classifier – 0.63
  - Gradient Boosting Classifier – 0.53
  - Naïve Bayes Classifier – 0.35
  - Multi-layer Perceptron – 0.61
  - Neural Network – 0.33
- Only the Random Forest Classifier truly performed well (relative to the others)
- For further experiments, we will consider only this type of classifier

# Results

- TF-IDF features extracted from raw text
  - Random Forest Classifier – 0.66
  - Best performing model overall
- TF-IDF features with 1- and 2-grams
  - Random Forest Classifier – 0.66
  - Same results as the previous model
- TF-IDF + PCA
  - Random Forest Classifier – 0.49
  - PCA gave only one feature for 95% of the variance, did not work well

# Results

- BERT embeddings from raw text, with a Random Forest Classifier
  - Only comment embeddings – 0.64
  - Comment embeddings and characteristics – 0.64
  - Comment embeddings, characteristics and parent embeddings – 0.63
- Surprisingly, the third method was the worst
- Cause – too many features or overfitting
- Could be solved using more training data and feature selection
  - Computationally intensive

# Results

- Interpretability
- Comment author and subreddit very important features
  - Person may be sarcastic all the time
  - Subreddit may require satirical content to be posted
- Some words and word pairs also important
  - “Yeah right!”
  - “Totally”
- Good sign as these words are also used in normal conversations

# Conclusion

- Random Forest, TF-IDF features from raw text – performed the best
- Our work was limited by computational constraints
- Despite using less data and simple methods, we came close to SOTA (0.71)
- Good embeddings and models improve complex task performance
- Future directions
  - Using more powerful computers, training on all available data
  - Fine-tune BERT and interpret it
  - Exploring other models or network architectures

# References


- Pushpak Bhattacharyya, *"Sarcasm Detection: A Computational and Cognitive Study."* Stanford University, CA, January 2018.
- Mikhail Khodak, Nikunj Saunshi, Kiran Vodrahalli, *"A Large Self-Annotated Corpus for Sarcasm."* arXiv (cs.CL), March 2018.
- Dan Ofer, *"Sarcasm on Reddit."* Kaggle, May 2018.


# Credits

- The Reddit logo is property of Reddit, Inc. and has been used here for representational purposes only.

# Q & A