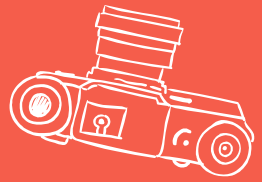
r/w266finalprojectspring2021



Detecting Sarcasm In Reddit Comments

By: Lavanya Vijayan & Shalini Kunapuli

Introduction

- Sarcasm: “A form of ironic speech intended to make fun of or criticize someone or something”
- Detecting sarcasm is a challenging NLP task
  - Several factors that shape the meaning of a phrase: tone, context, and history between two people conversing
- Prior work has been done in sarcasm detection via supervised learning
  - E.g. Lydia Xu and Vera Xu from Stanford used LSTM and BERT models to improve the accuracy of sarcasm detection by focusing on contextual information
- Our motivation
  - Build upon prior work, using BERT models and taking an unsupervised approach to improve sarcasm detection via cloze questions

Data

- Self Annotated Reddit Corpus (SARC)
- 1.3 million rows (one row per comment)
- Balanced in label (505413 sarc examples & 505413 non-sarc examples)
- Main columns: **parent_comment**, **comment**, **subreddit**, **upvotes**, **downvotes**, **date**, **label**
- Note: Here, **parent_comment** is a top comment or previous comment, and a **comment** is a response to it

parent_coment	comment	subreddit	upvotes	downvotes	date	label
...	...	...	...	...	...	...

Baseline

Convolutional Neural Network (CNN) using **comment** and **label** columns to train with two dense layers: ReLu and sigmoid

- Training accuracy: **0.7319**
- Testing accuracy: **0.7068**

Modeling

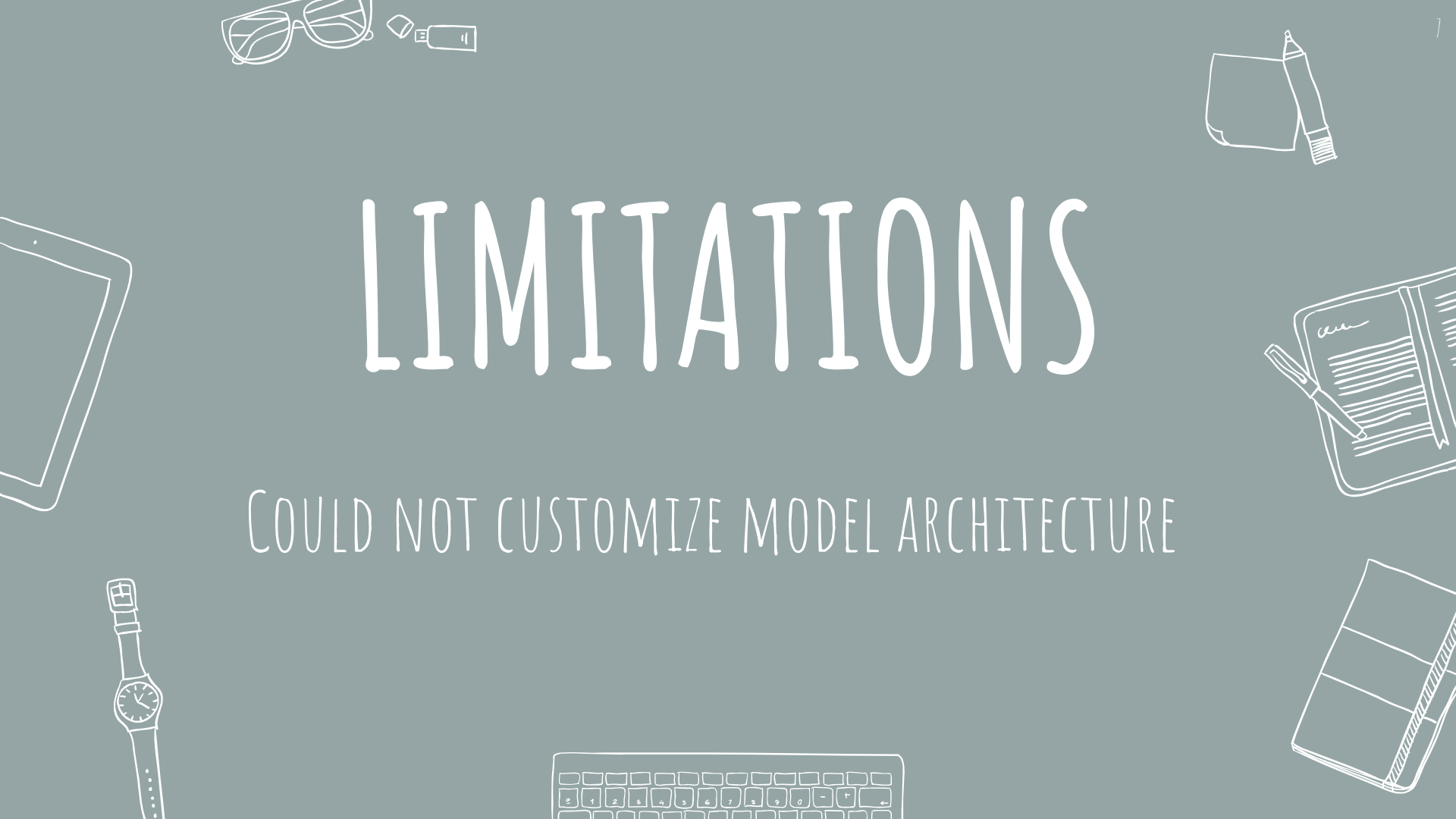
BERT, DistilBERT, Cloze + BERT

BERT

- Used **TFBertForSequenceClassification** to build a BERT model with a binary classification layer on top
- Used **TFTrainer** to train, evaluate, and predict
- Train on different amounts of balanced data, on both comment data and parent + comment data
  - [CLS][comment][SEP][PAD]
  - [CLS][parent][SEP][comment][SEP][PAD]
- Increasing the learning rate helped improve accuracy and loss

LIMITATIONS

Could not customize model architecture



DistilBERT

- Smaller, faster, lighter Transformer model
- Tokenized/Modeled on comment as well as parent + comment using **DistilBERTTokenizerFast**
  - [CLS][comment][SEP][PAD]
  - [CLS][parent][SEP][comment][SEP][PAD]
- Train with **TFDistilBertModel**
  - Various hyperparameters
  - Some models trained on freezed layers and others trained on unfreezed layers
- Training on unfreezed layers took longer but led to better accuracy

UNSUPERVISED

Applying the idea of cloze questions to few shot text classification

Cloze + BERT

Sarcastic related words	Non-sarcastic related words
joking	serious
ironic	honest
satirical	blunt
exaggerating	transparent
sarcastic	unsarcastic
kidding	

Test Phrases
"I am [MASK]!"
"I am being [MASK]!"
"OP is [MASK]!"
"OP is [MASK]."

"Fox News: Firefox most popular search engine after Microsoft and Apple"
"Foxnews is the most accurate reporting service next to the bible. I am [MASK]!"
Label is: 1 (sarcastic)
Logits:
kidding: **4.987739**
serious: 4.314688

• • •

"How old are you if I may ask?"
"I am 17, basically just turned. I am [MASK]!"
Label is: 0 (non-sarcastic)
Logits:
satirical: -3.0938523
honest: **0.4922531**

Cloze + BERT

- Sampled balanced datasets with **100, 200, 500,** and **1000** examples each
- For each sample, after finding the test phrase and pairs of words that most accurately classified the comments
  - Took an **average** of the logits and created soft labels for each comment; then ran through BERT

Cloze + BERT

Comparing soft labels and actual annotated labels

	Precision	Recall	F1 score	Accuracy
not sarc	0.52	0.72	0.61	0.53
sarc	0.55	0.34	0.42	0.53

Table II: Soft vs Actual Labels for 100 examples (50 per class)

	Precision	Recall	F1 score	Accuracy
not sarc	0.50	0.64	0.56	0.51
sarc	0.51	0.37	0.43	0.51

Table III: Soft vs Actual Labels for 200 examples (100 per class)

	Precision	Recall	F1 score	Accuracy
not sarc	0.50	0.68	0.57	0.50
sarc	0.49	0.31	0.38	0.50

Table IV: Soft vs Actual Labels for 500 examples (250 per class)

	Precision	Recall	F1 score	Accuracy
not sarc	0.50	0.69	0.58	0.50
sarc	0.49	0.30	0.37	0.50

Table V: Soft vs Actual Labels for 1000 examples (500 per class)

RESULTS

BERT, DistilBERT, Cloze + BERT

(With Comparison To Baseline)

Results

Hyperparameters	No. Train Examples	No. Val Examples	No. Test Examples	Accuracy	Loss
Training batch size 16, Evaluation batch size 64, Learning rate 5e-5, No. of Epochs 2, Weight decay 0.01, Hidden dropout prob 0.1	7,000	1,500	1,500	0.7910	0.4483
Training batch size 16, Evaluation batch size 64, Learning rate 5e-5, No. of Epochs 2, Weight decay 0.01, Hidden dropout prob 0.1	101,077	101,077	808,618	0.7613	0.4892

Table VI: BERT Model Results for only Comment Data

Hyperparameters	No. Train Examples	No. Val Examples	No. Test Examples	Accuracy	Loss
Training batch size 16, Evaluation batch size 64, Learning rate 5e-5, No. of Epochs 2, Weight decay 0.01, Hidden dropout prob 0.1	101,077	101,077	808,618	0.76217	0.48771

Table VII: BERT Model Results for Parent Comment + Comment Data

Model Architecture	No. Train Examples	No. Val Examples	No. Test Examples	Accuracy	Loss
Default Model	707,541	151,615	151,615	0.6537	0.6214
Additional Model Layers	707,541	151,615	151,615	0.8423	0.3594

Table VIII: DistilBERT Model Results for only Comment Data

Training Method	No. Train Examples	No. Val Examples	No. Test Examples	Accuracy	Loss
Freeze Layers for Training	707,541	151,615	151,615	0.6218	0.6489
Unfreeze Layers for Training	707,541	151,615	151,615	0.9299	0.1752

Table IX: DistilBERT Model Results for Parent Comment + Comment Data

Results Cont.

Data Size	Tokenizing	Test Phrase	Word Pairs	Accuracy	Loss
100 examples (50 per class)	Comment + Parent Comment	" I am being [MASK]!"	(('joking', 'transparent'), ('kidding', 'transparent'), ('joking', 'blunt'), ('ironic', 'transparent'), ('ironic', 'blunt'))	0.6406	0.6901
200 total (100 per class)	Comment + Parent Comment	" I am being [MASK]!"	(('joking', 'transparent'), ('kidding', 'transparent'), ('joking', 'blunt'), ('ironic', 'transparent'), ('ironic', 'blunt'))	0.5156	0.7348
500 total (250 per class)	Comment + Parent Comment	" I am being [MASK]!"	(('joking', 'transparent'), ('kidding', 'transparent'), ('sarcastic', 'serious'), ('ironic', 'transparent'), ('joking', 'blunt'))	0.6250	0.6815
1000 total (500 per class)	Comment + Parent Comment	" I am being [MASK]!"	(('joking', 'transparent'), ('kidding', 'transparent'), ('joking', 'blunt'), ('sarcastic', 'serious'), ('kidding', 'blunt'))	0.7266	0.5999

Table X: Cloze + BERT Model Results for Different Data Sizes

ERROR ANALYSIS

BERT, DistilBERT, Cloze + BERT

Error Analysis

Model	Parent Comment	Comment	Actual Label	Predicted Label
BERT	Man that's alot of bird poop	Don't count on it being cleaned up either!	sarcastic	non- sarcastic
BERT	What are some tips to help out a socially awkward person?	When I make eye contact with someone, I feel like it's a just a game to see who looks away first.	non- sarcastic	sarcastic
Distil BERT	i'd say avatar was one of the few exceptions, but that's probably just me	yep, it's only you.	sarcastic	non- sarcastic
Distil BERT	why are there records so cheap? lol	herb alpert sold something like 70 million records so you're bound to find em	non- sarcastic	sarcastic
Cloze + BERT	Local paper does piece on Grandma Shirley	Wow, that is awesome.	sarcastic	non- sarcastic
Cloze + BERT	Bernicrat. Ugh	Y'know,the people who registered as a Democrat two weeks ago and call everyone who disagrees with them DINOs.	non- sarcastic	sarcastic

Table XI: Example False Positives and False Negatives per Model

Error Analysis Cont.

- Preface: Subjectivity of sarcasm could lead the labels to be different if a different set of people annotated them in the first place → Errors and our analysis of the errors are conditional on those who self-annotated this dataset
- Slang including Internet slang are used to express a variety of feelings
  - E.g. "lol" and "ugh" can be used in sarcastic and non-sarcastic expressions
    - Model misclassified non-sarcastic examples containing these words
    - While communicating digitally, p eople often throw "lol" into phrases even when they are talking about something serious → Model might not be accounting for this

		Actual Label	Predicted Label
why are there records so cheap? lol	herb alpert sold something like 70 million records so you're bound to find em	non- sarcastic	sarcastic

		Actual Label	Predicted Label
Bernicrat. Ugh	Y'know,the people who registered as a Democrat two weeks ago and call ev-eryone who disagrees with them DINOs.	non- sarcastic	sarcastic

Error Analysis Cont.

- Another misclassified example contains “wow,” which is intended sarcastically, but the model interprets it as non-sarcastic
  - “wow” has a duality; it can be used by someone who is genuinely surprised, unsurprised, impressed, or unimpressed, to name a few
  - Understanding these nuances is something people develop over time through diverse social interactions
  - Difficult to capture these nuances in a way that can be taught to the model

		Actual Label	Predicted Label
Local paper does piece on Grandma Shirley	Wow, that is awesome.	sarcastic	non- sarcastic

- ”Don’t count on it” is a phrase that can be said jokingly or seriously
- Within the topic of ”bird poop,” the phrase is meant as a joke; the sarcasm is delivered through the tone rather than the literal meaning of the phrase
  - Model might have interpreted it as non-sarcastic based on ”count” and ”cleaned,” which have empirical connotations and could be found in many serious responses.

		Actual Label	Predicted Label
Man that’s alot of bird poop	Don’t count on it being cleaned up either!	sarcastic	non- sarcastic

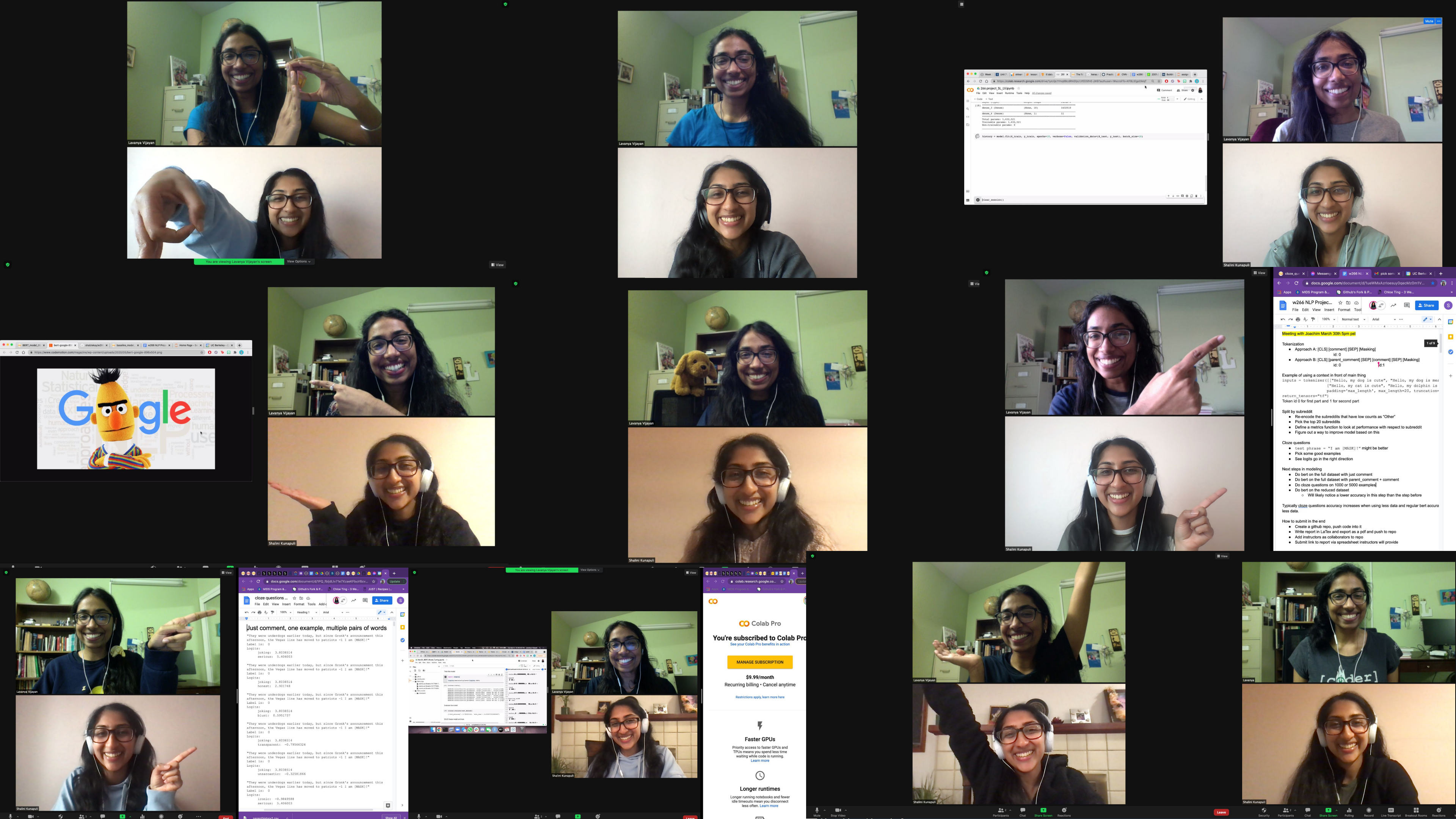
Error Analysis Cont.

		Actual Label	Predicted Label
What are some tips to help out a socially awkward person?	When I make eye contact with someone, I feel like it's a just a game to see who looks away first.	non- sarcastic	sarcastic

- Author of parent comment seems to be genuinely asking fellow Redditors for advice
- Author of the comment seems to share a strategy that helps them face social situations in an honest and humorous way
- Model interprets the advice as sarcastic, which is likely due to the presence of the word ”game,” which has a playful connotation
- Model likely sees “game” in a non-serious context and “tips” in a serious context, via prior data; Also model might see “game” more often than “tips” → This could lead model to have higher propensity for “sarcastic” in cases like this where both these words are present

Conclusion & Future Work

- Best model accuracies
  - BERT: **0.7622**
  - DistilBERT: **0.9299**
  - Cloze + BERT: **0.7266**
- Supervised generates better results as compared to unsupervised
  - Sarcasm is highly subjective
  - Supervised performs better due to labels determined by human annotators
- In the future
  - Add more context related to subreddit
  - Generation of sarcastic response based on a previous phrase


Lavanya Vijayan
Shalini Kunapuli
You are viewing Lavanya Vijayan's screen
View Options
w266 NLP Projec...
Meeting with Joachim March 30th 5pm pst
Tokenization
Approach A: [CLS] [comment] [SEP] [Masking]
Approach B: [CLS] [parent_comment] [SEP] [comment] [SEP] [Masking]
Example of using a context in front of main thing
Token id 0 for first part and 1 for second part
Split by subreddit
Re-encode the subreddits that have low counts as "Other"
Pick the top 20 subreddits
Define a metrics function to look at performance with respect to subreddit
Figure out a way to improve model based on this
Cloze questions
Pick some good examples
See logits go in the right direction
Next steps in modeling
Do bert on the full dataset with just comment
Do bert on the full dataset with parent_comment + comment
Do cloze questions on 1000 or 5000 examples
Do bert on the reduced dataset
Will likely notice a lower accuracy in this step than the step before
How to submit in the end
Create a github repo, push code into it
Write report in LaTex and export as a pdf and push to repo
Add instructors as collaborators to repo
Submit link to report via spreadsheet instructors will provide
cloze questions
Just comment, one example, multiple pairs of words
Colab Pro
You're subscribed to Colab Pro
See your Colab Pro benefits in action
MANAGE SUBSCRIPTION
$9.99/month
Recurring billing • Cancel anytime
Restrictions apply, learn more here
Faster GPUs
Longer runtimes
Leave


ANY QUESTIONS?
MacBook Air

Slides Carnival

Free templates for all your presentation needs

For PowerPoint and Google Slides

100% free for personal or commercial use

Ready to use, professional and customizable

Blow your audience away with attractive visuals