
NLP Project - Human Comment Generator

Team #20

Duo Jiang(dj1057)

Weilong Chen(wc1660)

Ziyi Xie(zx1153)

Yihang Zhang(yz2865)

Outline

- Objective
- Data
- Models
- Evaluation and Sample Result
- References

Objective

- Generate human-like replies that are likely to be upvoted given comments on Reddits
 - Explore the viability of doing so using deep learning models and transfer learning
 - Applications:
 - Explore the quality of comment text generated by model
 - Explore if difference between machine generated text and human generated text can be discerned by human
-

Data

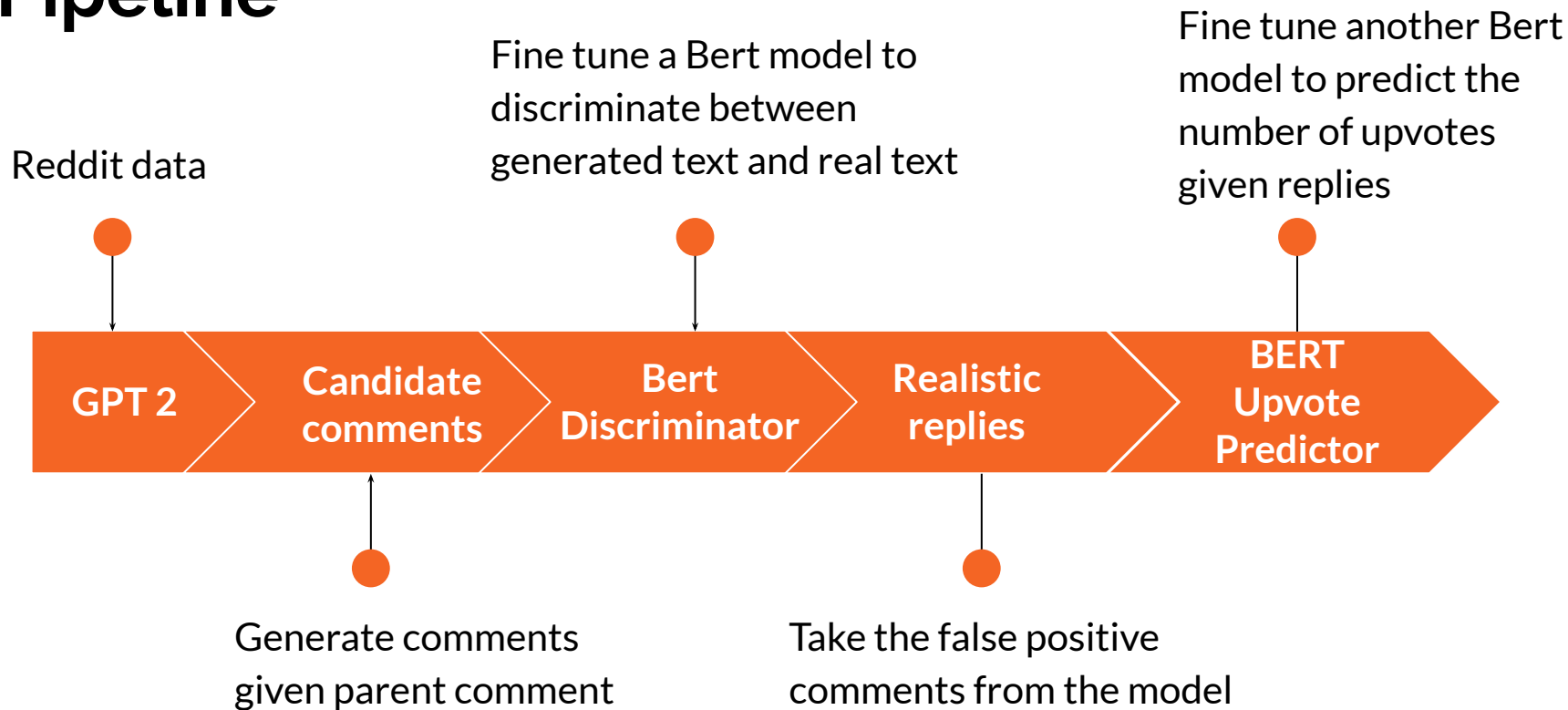
- **Kaggle**

- <https://www.kaggle.com/ehallmar/reddit-comment-score-prediction>
- Comments with positive/negative upvotes
- Important Attributes: id, parent_id, subreddit_id, text, scores, parent_link_id, parent_text, parent_score, parent_ups

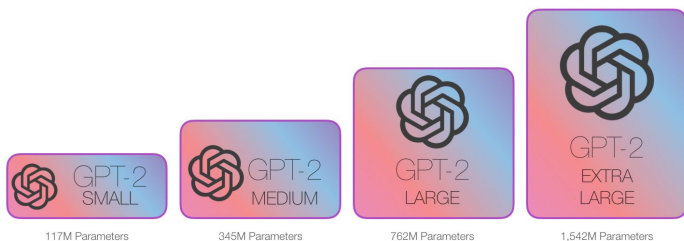
- **Pushshift**

- <https://pushshift.io/> with Google Bigquery
 - Collect data on specific topics with 'subreddit' attribute
-

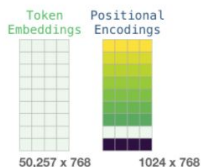
Pipeline



Pre-trained GPT2 model and Fine tuning



GPT-2 SMALL

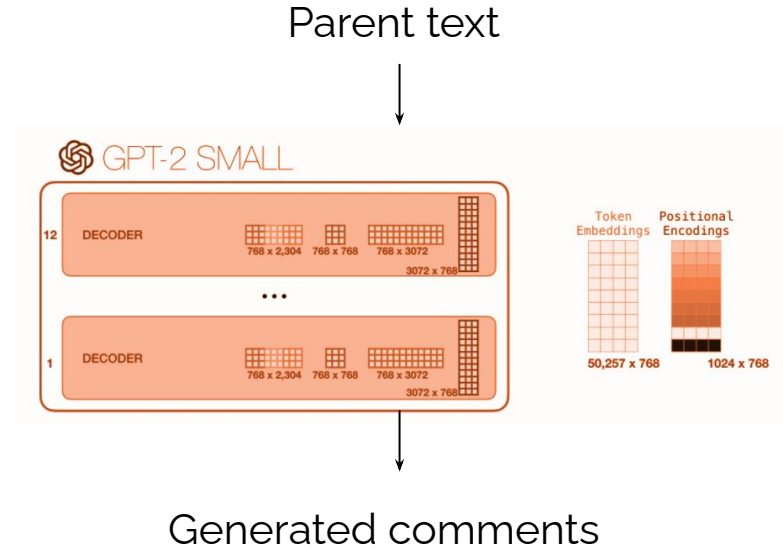
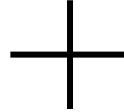


text	parent_text
Either way, you know something is coming.	Difference between foreshadowing and foreplay....
I learned that the hard way, He's sitting over...	Never assume it's safe to cum inside her.
Now that's a...high dive.	Stemless high quality weed.\n\nsoft landing. ...
Yes! Absolutely. I speak Russian to the baby a...	Do you plan on raising your children to speak ...
Whenever there's a fire alarm at my work, we a...	Wooow, if there was a fire drill at my work th...
Leonardo Dicaprio will be safe.	Would you rather fight one 125' Jesus in Rio o...
these keeeeds	How do I reach these kids?!
Was it Ranjit? Everyone knows that motherfucker.	I sat in an interview with my boss who is alwa...
HIPPAW forbids it.	Doctor/Patient confidentiality.

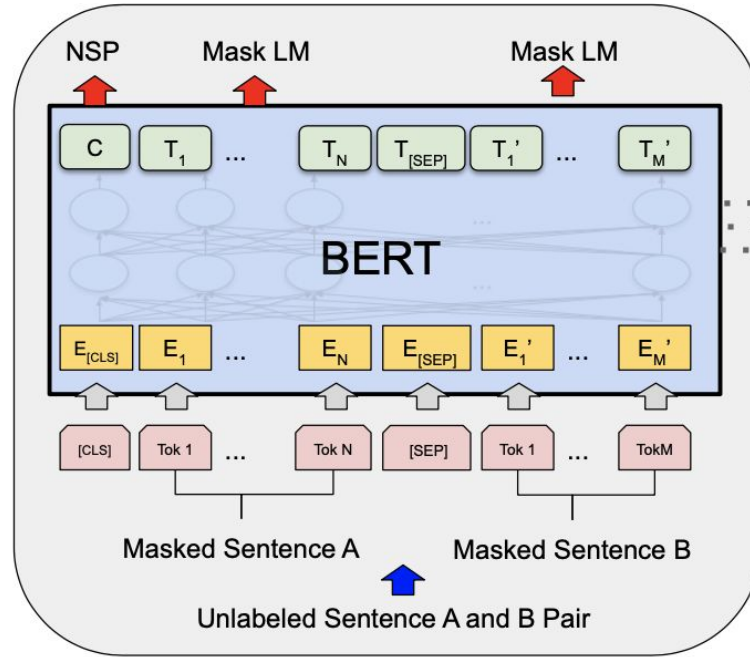
72h to train
6 validation complexity

Text Generation after Fine Tuning

- ❑ Max Length
- ❑ Top p
- ❑ Top k
- ❑ Temperature



Bert architecture



Pre-training

We use Bert for two purposes

1. Classification
(Exclude replies GPT generate that can be noise)
2. Regression
(Extract replies with high upvotes)

Bert-Realistic Discriminator

- Aim to predict the probability of being a real reply.
 - Label: '1' as Real, '0' as Fake.
 - Trained by real comments data + 'fake' comments data generated by GPT2 model
-

Bert - Upvote Predictor

- Aim to predict the if a reply will get upvotes or not.
 - Serve as another filter for select comments of high quality
 - Trained by real reddit comments data with # of upvotes
-

Result - Sample Comments

Comment:

In hindsight, I have no idea why I thought this would work. But, science was done.

Human replies



Your debacle was quite the spectacle.

Generated replies



1. A true master of this thread. We don't even think in russia
2. I thought you said your dad would know because... science... nevermind.

Evaluation

- Bert performance on discriminating real comments vs. generated comments
 - Quality of generated comments:
 - We use human evaluation result
-

Evaluation

Bert-Realistic Discriminator

Performance

- 1400 real Reddit data + 'fake' data generated by GPT2 (comments + replies)
- ACC: 0.96
- TP: 696
- TN: 681
- FP: 17
- FN: 6

Get the Best Fake

- 4605 'fake' data generated by GPT2
 - About 90 replies for each comment
 - ACC: 0.91
 - TN: 4193
 - FP: 412
 - 412 FP replies would be fed into the Bert-Upvote for evaluation
-

Poll Form

1. Evaluate if human can discern machine generated comments from human comments
2. Evaluate the effectiveness of the Bert Upvote Predictor

Q1 part 1: Mexico is the real papá. US kidnapped adopted Texas. Along with California, Arizona, New Mexico... *

Please pick the true comment that you think is not generated by machine.

Nah, Texas ran aw... And all the way in t... "We'll help." - US pr... "MEXCIA! MELIZA...

The True Comment ☐ ☐ ☐ ☐

Q1 Part 2: Mexico is the real papá. US kidnapped adopted Texas. Along with California, Arizona, New Mexico... *

Please order the comments according to your willingness to upvote them

"MEXCIA! MELIZA... "We'll help." - US pr... Nah, Texas ran aw... And all the way in t...

Upvote Order: First	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Upvote Order: Seco...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Upvote Order: Third	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Upvote Order: Fourth	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

The poll results are still in collection process and will be included in the final report.

References

GPT2:

- DialoGPT: Large-Scale Generative Pre-training for Conversational Response Generation,author={Yizhe Zhang and Siqi Sun and Michel Galley and Yen-Chun Chen and Chris Brockett and Xiang Gao and Jianfeng Gao and Jingjing Liu and Bill Dolan},year={2020}

BERT:

- BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (<https://arxiv.org/abs/1810.04805>)
-

Next steps

Assignment 1

Lorem ipsum dolor sit amet

Consectetur adipiscing elit, sed do eiusmod tempor

Assignment 2

Lorem ipsum dolor sit amet

Assignment 3

Consectetur adipiscing elit, sed do eiusmod tempor
