Microsoft Research AI

Dialogue Response Ranking Training with Large-Scale Human Feedback Data

Xiang Gao, Yizhe Zhang, Michel Galley, Chris Brockett, Bill Dolan
Microsoft Research AI, Redmond, WA, USA

paper: arxiv.org/abs/2009.06978
code: [github.com/golsun/**DialogRPT**](github.com/golsun/DialogRPT)
data: https://dialogfeedback.github.io

Motivation

- Great progress in building conversational AI with large-scale pre-trained models
- They are trained mostly by minimizing **perplexity** on human samples

Microsoft

DialoGPT: arxiv.org/abs/1911.00536
Trained with **147 M** Reddit Dialogues!

Google Brain

Meena: arxiv.org/abs/2001.09977

FACEBOOK

Blender: arxiv.org/abs/2004.13637

And many other awesome works..

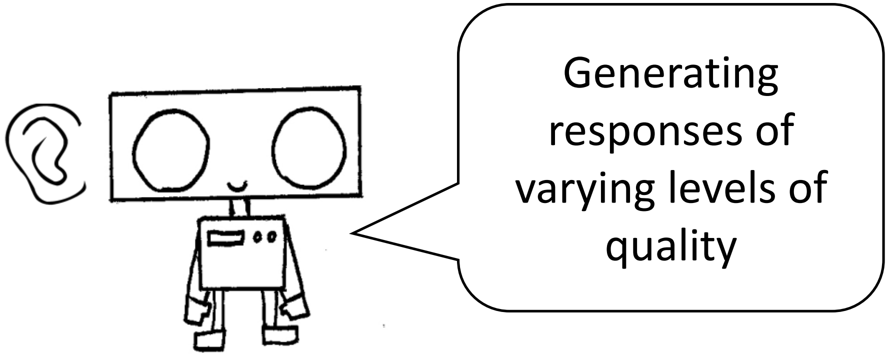
Motivation

- Great progress in building conversational AI with large-scale pre-trained models
- They are trained mostly by minimizing **perplexity** on human samples
- However, some human replies are more engaging than others, spawning more follow-up interactions

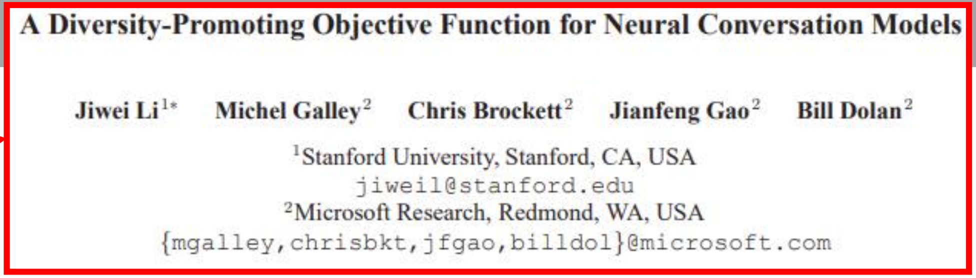
engaging

Boring/bland/generic

Hate/offensive/toxic



Motivation


A Diversity-Promoting Objective Function for Neural Conversation Models

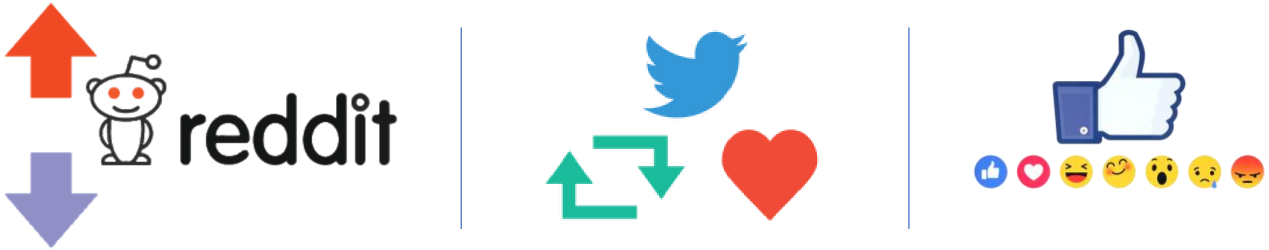
Jiwei Li[1*] Michel Galley[2] Chris Brockett[2] Jianfeng Gao[2] Bill Dolan[2]

[1]Stanford University, Stanford, CA, USA
jiweil@stanford.edu
[2]Microsoft Research, Redmond, WA, USA
{mgalley,chrisbkt,jfgao,billdol}@microsoft.com

- Existing ranking methods may be suboptimal
  - **Perplexity**: e.g. MMI, only reflects relevancy
  - **Manually designed features**: not directly based on real-world human preferences in an end-to-end fashion.

- Crowdsourcing of large-scale training data is too expensive

- Social networks provide ways to measure Human feedback on dialogues (and other contents).



Motivation

- Optimizing expected ***human feedback***, not just ***perplexity***?
- Social network human feedback data!

Microsoft Research AI

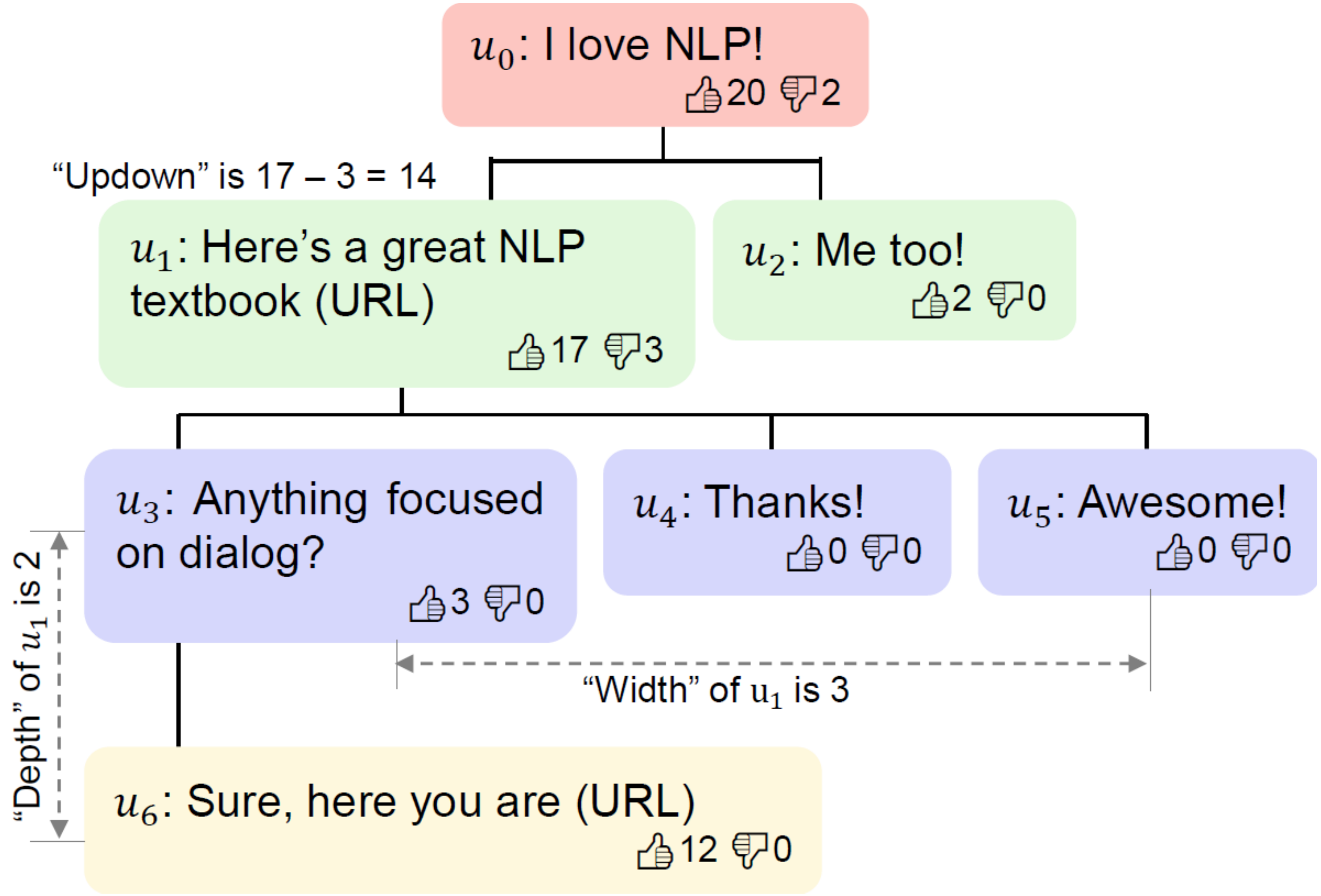
- Motivation
- **Dataset**
- Method
- Results

Dialogue Response Ranking Training with Large-Scale Human Feedback Data
Xiang Gao, Yizhe Zhang, Michel Galley, Chris Brockett, Bill Dolan
Microsoft Research AI, Redmond, WA, USA

paper: arxiv.org/abs/2009.06978
code: github.com/golsun/**DialogRPT**
data: https://dialogfeedback.github.io

EMNLP 2020
The 2020 Conference on Empirical Methods in Natural Language Processing
16th – 20th November 2020

Dataset

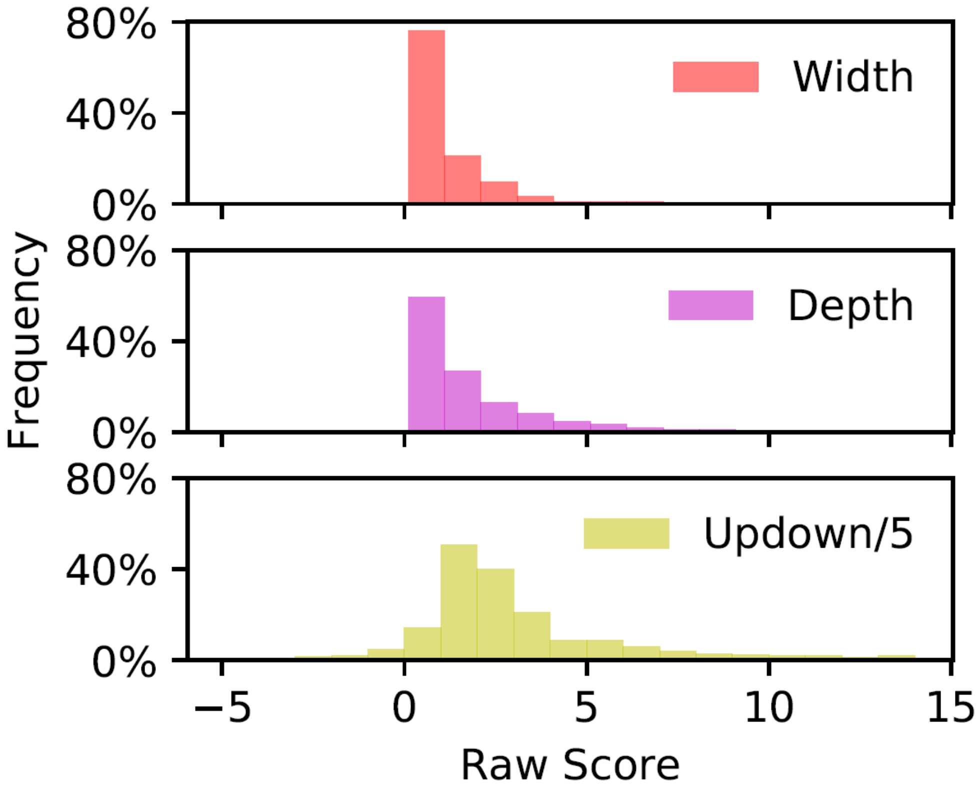


- We define three metrics of human feedback on Reddit
  - Updown
  - Width
  - Depth

Dataset

Microsoft Research AI

- All of three metrics have a long-tailed distribution



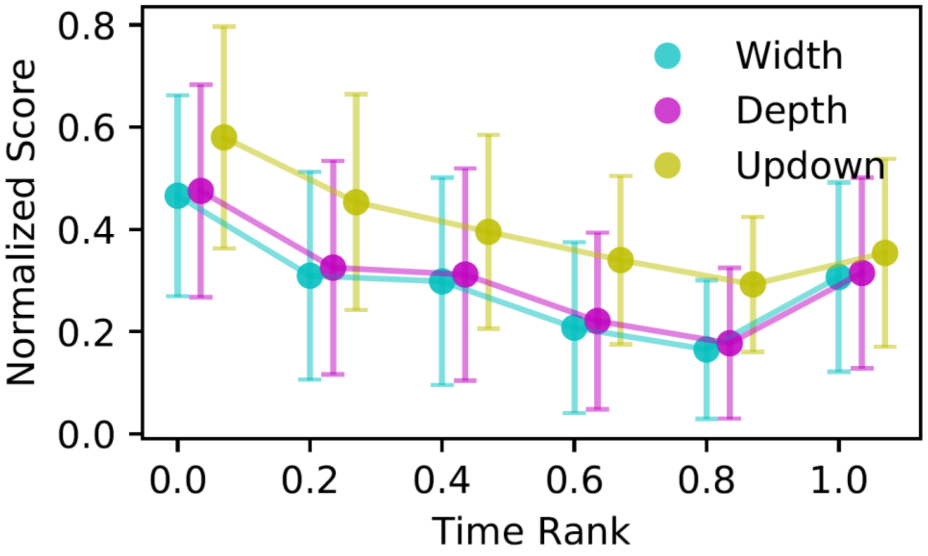
- They are correlated. Width and depth are more correlated as both are measure of number of replies

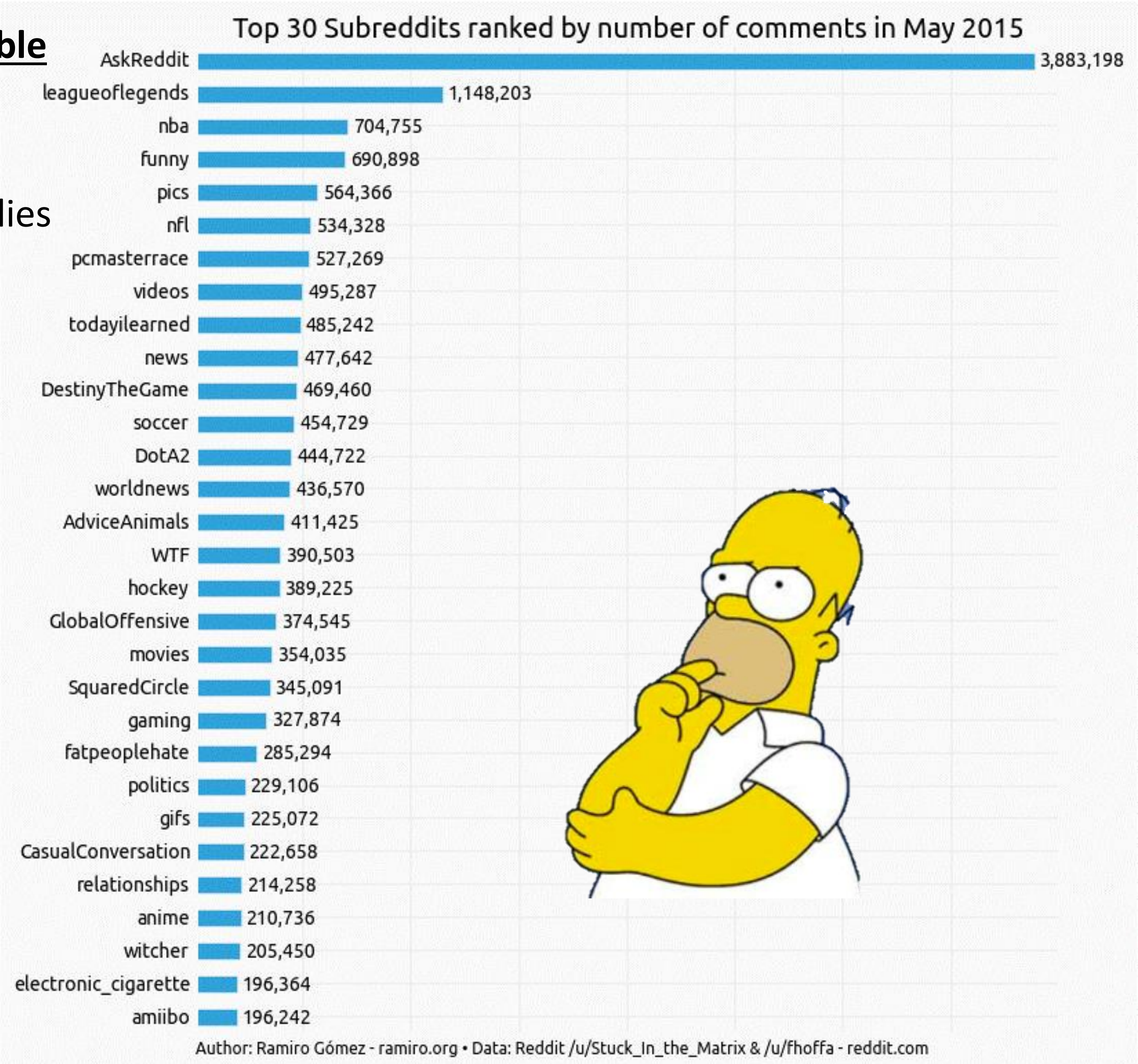
	Width	Depth	Updown
Width	1	0.8592	0.3491
Depth	0.8592	1	0.3257
Updown	0.3491	0.3257	1

Table 1: Spearman's ρ between different measurements of human feedback. Darker cell color indicates higher correlation.

Dataset

- However, these metrics are **not directly usable/comparable**
- Confounding factors
  - **Topics/Subreddits**: popularity differs significantly
  - **Timing**: The early comments gets more upvotes/replies





paper: arxiv.org/abs/2009.06978
code: [github.com/golsun/**DialogRPT**](https://github.com/golsun/DialogRPT)
data: https://dialogfeedback.github.io

Microsoft Research AI

- Motivation
- Dataset
- **Method**
- Results

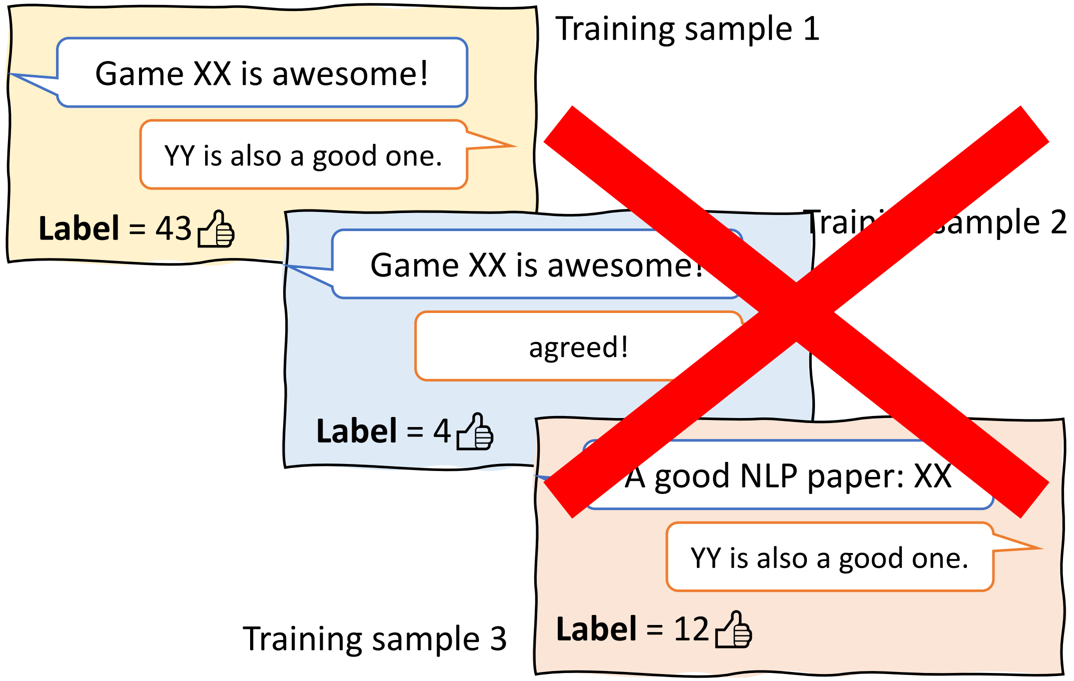
Xiang Gao, Yizhe Zhang, Michel Galley, Chris Brockett, Bill Dolan
Microsoft Research AI, Redmond, WA, USA

Contrastive Dataset

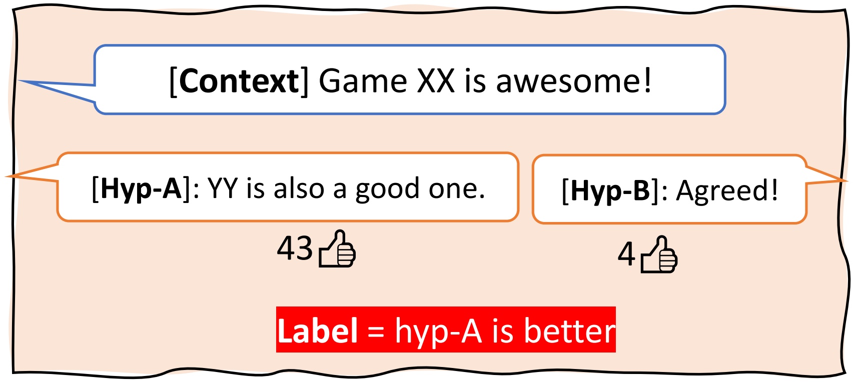
Microsoft Research AI

- Directly predicting metric value is **hard**, due to confounding factors (e.g. timing of post)
- **Contrastive learning!**

Predicting feedback metric value



Classify which one gets more feedback



Only compare pairs of responses that are comparable

- For the same dialogue context
- Published at roughly the same time
- …

paper: arxiv.org/abs/2009.06978
code: github.com/golsun/DialogRPT
data: https://dialogfeedback.github.io

Contrastive Learning

Microsoft Research AI

h = **Logits** (scalar)

c = **Context** (string)

$$h(c, r) = \textsc{DialogRPT}(c, r)$$

r = **Response** (string)

- Inferred Score

$$s(r|c) = \text{Sigmoid}(h(c, r))$$

- Training Loss

$$\mathcal{L} = -\sum_{i \in \text{batch}} \log \boxed{\frac{e^{h(c_i, r_i^+)}}{e^{h(c_i, r_i^+)} + e^{h(c_i, r_i^-)}}}$$

negative log likelihood

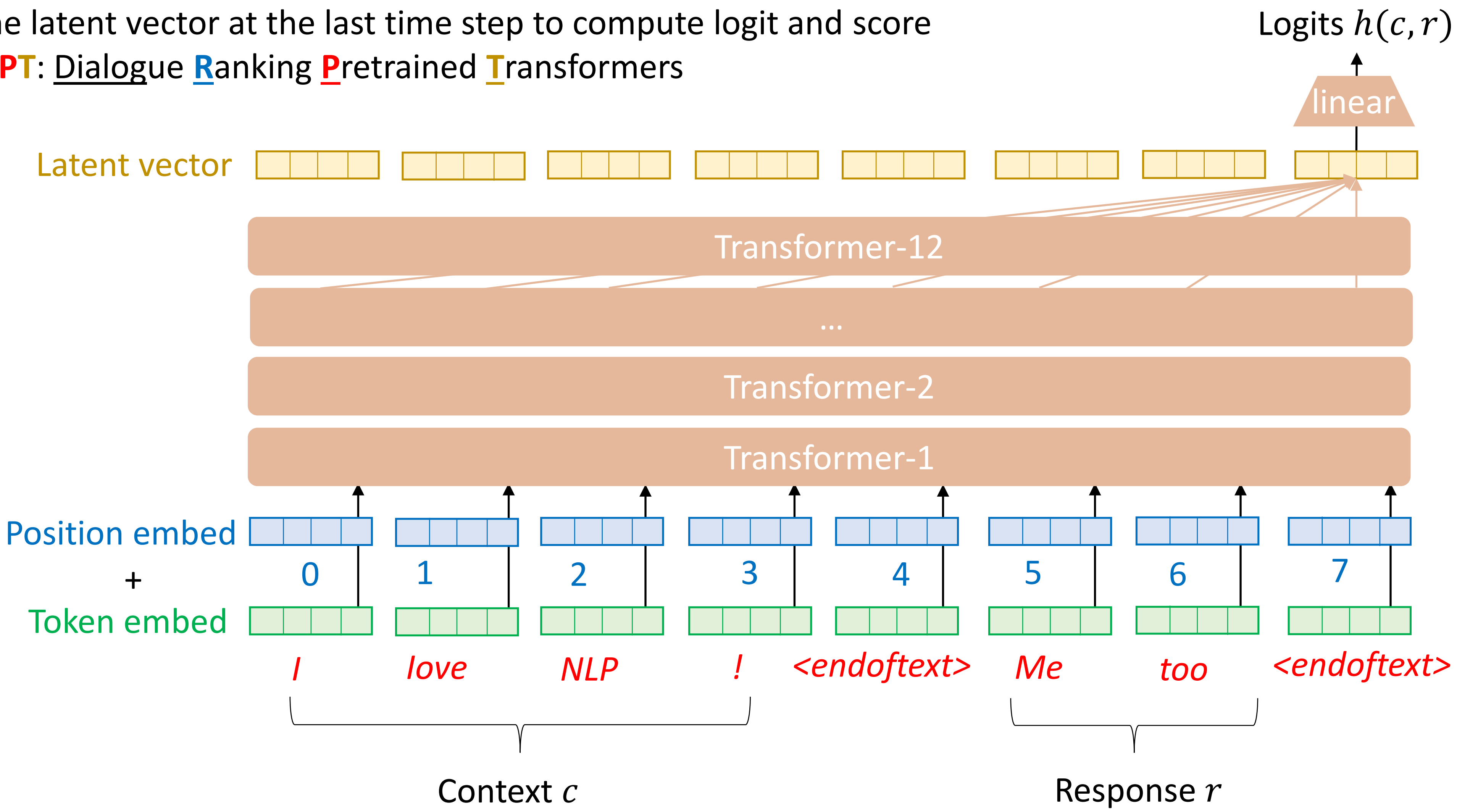
= $P(r^+|c)$

Softmax probability to pick r^+

- r^+ for positive sample
- r^- for negative sample

Implementation

- GPT-2 type model, initialized with DialoGPT weight
- Using the latent vector at the last time step to compute logit and score
- **DialogRPT**: Dialogue Ranking Pretrained Transformers



paper: arxiv.org/abs/2009.06978
code: github.com/golsun/**DialogRPT**
data: https://dialogfeedback.github.io

Model ensemble

Microsoft Research AI

However, can we apply rankers trained on human vs human data on generators?

r is a human-like response: $r \in H$

Probability that response r gets the most feedback given context c:

=0, assumed

$$P(r|c) = P(r|c, r \in H)P(r \in H) + P(r|c, r \notin H)P(r \notin H)$$

$$= P(r|c, r \in H)P(r \in H)$$

Task	Subtask Description	Training size (number of pairs)
$P(r \in H)$ **Human vs fake**	Fake = Retrieved human response	40.7 M
	Fake = Machine generated response	40.7 M
$P(r\|c, r \in H)$ **Human vs human** (which gets more feedback)	Feedback = Updown (more upvotes - downvotes)	40.7 M
	Feedback = Width (more direct replies)	22.3 M
	Feedback = Depth (longer follow-up thread)	25.1 M

paper: arxiv.org/abs/2009.06978
code: github.com/golsun/**DialogRPT**
data: https://dialogfeedback.github.io

Microsoft Research AI

- Motivation
- Dataset
- Method
- **Results**

Dialogue Response Ranking Training with Large-Scale Human Feedback Data
Xiang Gao, Yizhe Zhang, Michel Galley, Chris Brockett, Bill Dolan
Microsoft Research AI, Redmond, WA, USA

paper: arxiv.org/abs/2009.06978
code: github.com/golsun/**DialogRPT**
data: https://dialogfeedback.github.io

EMNLP 2020
The 2020 Conference on Empirical Methods in Natural Language Processing
16th – 20th November 2020

Example

Microsoft Research AI

Context: I love NLP!				
Response:		Width	Depth	Updown
A	Me too!	0.033	0.043	0.171
B	It's super useful and more and more powerful!	0.054	0.164	0.296
C	Can you tell me how it works?	0.644	**0.696**	0.348
D	Can anyone recommend a nice review paper?	**0.687**	0.562	0.332
E	Here's a free textbook (URL) in case anyone needs it.	0.319	0.409	**0.612**

Table 3: Predicted feedback scores of several example responses given the same context.

Generator reranking

Although hypothesis C is most likely to be generated (Generation Probability = 0.496), it’s relatively boring. Using Updown Score, we can pick the hypothesis A, which is perhaps more interesting (Updown Score = 0.431)

[Context]: Can we restart 2020?

	Generation Probability	Updown Score	Generated Hypothesis
A	0.383	0.431	I think we should go back to the beginning, and start from the beginning.
B	0.195	0.323	I think we should just give up and let the year just pass.
C	0.496	0.302	Yes, we can.
D	0.328	0.153	I think so, yes.

Human vs Human Performance

	Method	Pairwise accuracy	Spearman ρ
Width	Dialog ppl.	0.513	-0.009
	Reverse dialog ppl.	0.571	0.099
	Length baseline	0.595	0.229
	BoW baseline	0.596	0.234
	DIALOGRPT	**0.752**	**0.357**
Depth	Dialog ppl.	0.508	-0.004
	Reverse dialog ppl.	0.557	0.063
	Length baseline	0.543	0.134
	BoW baseline	0.584	0.187
	DIALOGRPT	**0.695**	**0.317**
Updown	Dialog ppl.	0.488	0.003
	Reverse dialog ppl.	0.560	0.076
	Length baseline	0.531	0.063
	BoW baseline	0.571	0.134
	DIALOGRPT	**0.683**	**0.295**

Table 5: Performance on test set ranking gold responses, measured by pairwise accuracy and Spearman's ρ.

Human vs Rand Performance

For each context, there're k human responses and n distractor (random human responses),
Rank these $k + n$ candidates based on predicted $P(r \in H)$
Even DialogRPT is only trained on Reddit, it performs very well on all four datasets

Dataset	Method	Hits@k
Reddit (k >5,n=k)	BLEU1	0.651
	BERTScore	0.685
	BLEURT	0.714
	BM25	0.309
	ConvRT	0.760
	Dialog ppl.	0.560
	Reverse dialog ppl.	0.775
	DIALOGRPT	**0.886**
DailyDialog (k=1,n=19)	BM25	0.182
	ConvRT	0.380
	Dialog ppl.	0.176
	Reverse dialog ppl.	0.457
	DIALOGRPT	**0.621**

Zero-short!

Dataset	Method	Hits@k
Twitter (k=1,n=19)	BM25	0.178
	ConvRT	0.439
	Dialog ppl.	0.107
	Reverse dialog ppl.	0.440
	DIALOGRPT	**0.548**
PersonaChat (k=1,n=19)	BM25	0.117
	ConvRT	0.197
	IR Baseline	0.213
	Starspace	0.318
	KV profile memory	0.349
	Dialog ppl.	0.108
	Reverse dialog ppl.	0.449
	DIALOGRPT	**0.479**

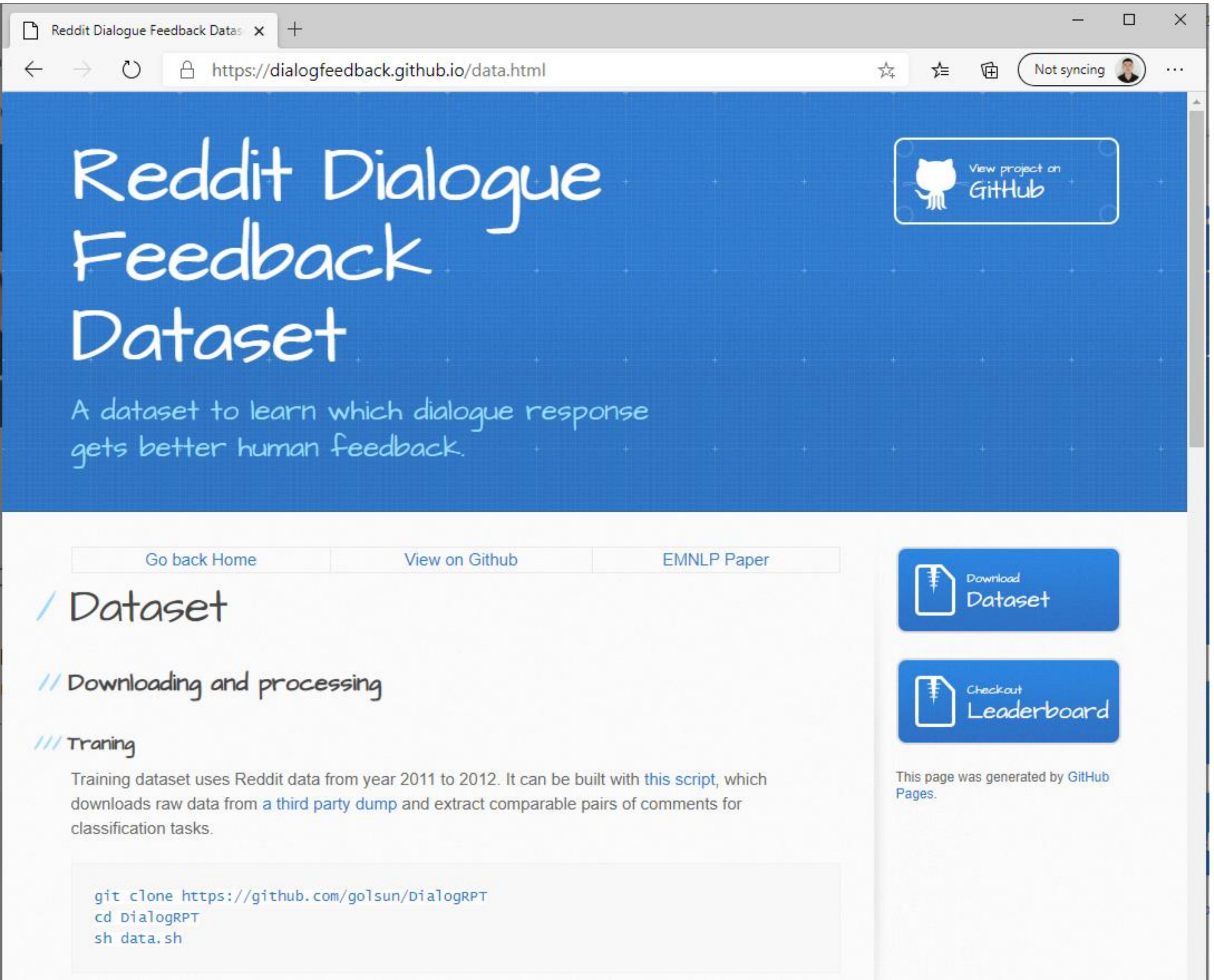
Ensembled model

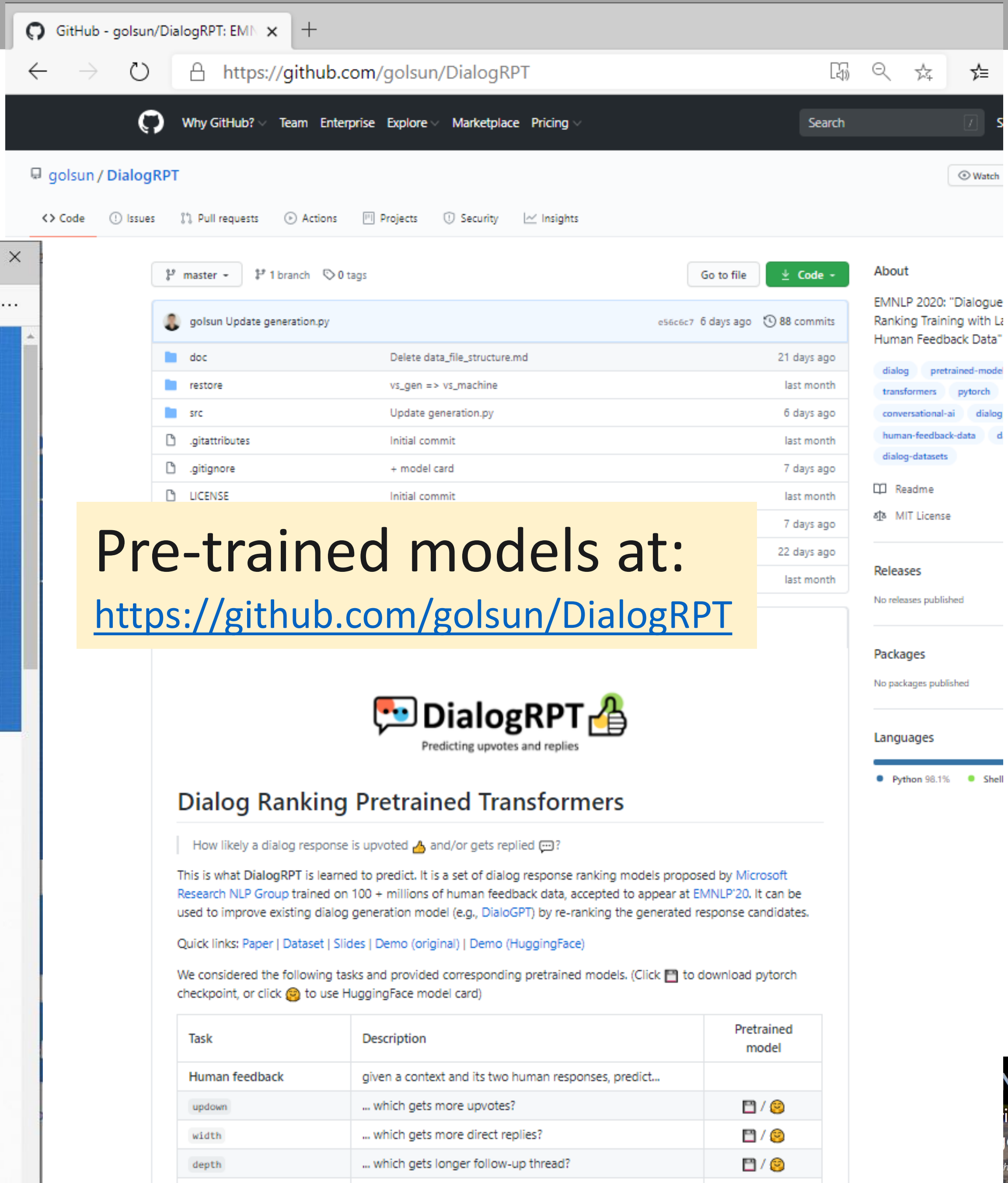
Model	Trained on	Tested on: Human vs. Human Width	Depth	Updown	Human vs. Fake Rand	Generated
Human feedback	Width	0.764	0.693	0.601	0.517	0.644
	Depth	0.749	0.701	0.588	0.512	0.647
	Updown	0.659	0.602	0.683	0.526	0.667
Human-like	Rand	0.558	0.552	0.522	0.843	0.413
	+ Generated	0.560	0.558	0.522	0.864	0.880
Ensemble	-	0.746	0.675	0.666	0.758	0.821

Table 6: Pairwise accuracy of DIALOGRPT models. Darker cell color indicates better performance.

Open-sourced!

Dataset available at:
https://dialogfeedback.github.io





Pre-trained models at:
https://github.com/golsun/DialogRPT

Microsoft Research AI

Open-sourced!

DialogRPT Demo (original).ipynb

https://colab.research.google.com/drive/1jQXzTYsgdZIQjJKrX4g3CP0_PGCeVU3C?usp=sharing

File Edit View Insert Runtime Tools Help Changes will not be saved

+ Code + Text Copy to Drive

```
!python src/score.py play -p=restore/updown.pth

100% 1042301/1042301 [00:00<00:00, 2691834.25B/s]
100% 456318/456318 [00:00<00:00, 1772148.58B/s]
--2020-09-16 00:17:24--  https://xiagnlp2.blob.core.windows.net/dialogrp
Resolving xiagnlp2.blob.core.windows.net (xiagnlp2.blob.core.windows.net
Connecting to xiagnlp2.blob.core.windows.net (xiagnlp2.blob.core.windows
HTTP request sent, awaiting response... 200 OK
Length: 1520029114 (1.4G) [application/octet-stream]
Saving to: 'restore/updown.pth'

                                1.42G  65.6MB/s    in 2
own.pth' saved [1520029114

use `_EOS_` to delimite turns for a multi-turn context

Context:  I love NLP!
Response: Here's a free textbook (URL) in case anyone needs it.
score = 0.613

Context:  I love NLP!
Response: Me too!
score = 0.111
```

Demo available at:
http://github.com/golsun/DialogRPT

DialogRPT Demo (Hugging Face).ipynb

https://colab.research.google.com/dr

File Edit View Insert Runtime Tools Help Cannot save changes

+ Code + Text Copy to Drive

Step 2. Set up the `DialogRPT` model using Hugging Face model card

```
from transformers import AutoTokenizer, AutoModelForSequenceClassification
import torch

model_card = "microsoft/DialogRPT-updown"   # you can try other model_card listed in the table above
tokenizer = AutoTokenizer.from_pretrained(model_card)
model = AutoModelForSequenceClassification.from_pretrained(model_card)

def score(cxt, hyp):
  model_input = tokenizer.encode(cxt + "<|endoftext|>" + hyp, return_tensors="pt")
  result = model(model_input, return_dict=True)
  return torch.sigmoid(result.logits)
```

Step 3. Play!

In the following example, the model predicts that, given the same context "I love NLP!", response B is gets more upvo

	Response of "I love NLP!"	Score
A	Me too!	0.111
B	Here's a free textbook (URL) in case anyone needs it.	0.613

```
cxt = "I love NLP!"
hyp_A = "Me too!"
hyp_B = "Here's a free textbook (URL) in case anyone needs it."

print('%.3f %s'%(score(cxt, hyp_A).squeeze(), hyp_A))
print('%.3f %s'%(score(cxt, hyp_B).squeeze(), hyp_B))

0.111 Me too!
0.613 Here's a free textbook (URL) in case anyone needs it.
```

HUGGING FACE

Back to all models

Model: `microsoft/DialogRPT-updown`

pytorch gpt2 text-classification pipeline:text-generation

paper: arxiv.org/abs/2009.06978
code: github.com/golsun/**DialogRPT**
data: https://dialogfeedback.github.io

Microsoft Research AI

Thank you!

Dialogue Response Ranking Training with Large-Scale Human Feedback Data
Xiang Gao, Yizhe Zhang, Michel Galley, Chris Brockett, Bill Dolan
Microsoft Research AI, Redmond, WA, USA

paper: arxiv.org/abs/2009.06978
code: github.com/golsun/**DialogRPT**
data: https://dialogfeedback.github.io