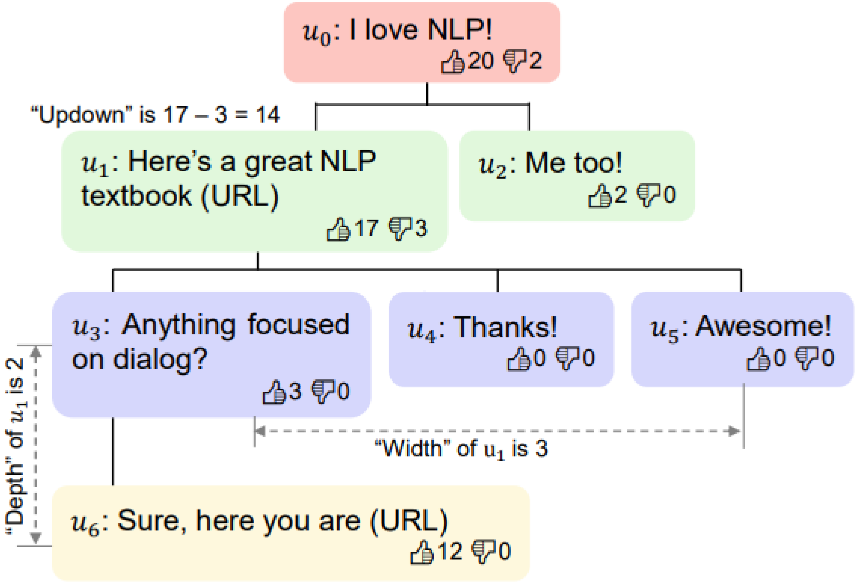
Dialogue Response Ranking Training with Large-Scale Human Feedback Data

X. Gao, Y. Zhang, M. Galley, C. Brockett, B. Dolan
Microsoft Research NLP

Long paper at EMNLP 2020
paper: arxiv.org/abs/2009.06978
code: github.com/golsun/DialogRPT
data: https://dialogfeedback.github.io

> “How likely a response gets upvoted?”
> -- Let’s optimize expected human feedback, instead of just perplexity.



u_0: I love NLP! 👍20 👎2

“Updown” is 17 – 3 = 14

u_1: Here’s a great NLP textbook (URL) 👍17 👎3

u_2: Me too! 👍2 👎0

u_3: Anything focused on dialog? 👍3 👎0

u_4: Thanks! 👍0 👎0

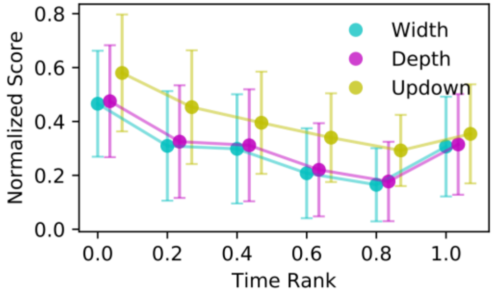
u_5: Awesome! 👍0 👎0

“Width” of u_1 is 3

“Depth” of u_1 is 2

u_6: Sure, here you are (URL) 👍12 👎0

Many confounding factors, e.g. timing and subreddit

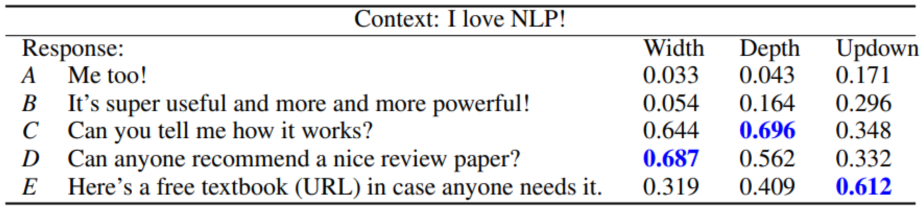


So instead of directly predicting scores, we train models to predict which one of a pair of “comparable” responses gets better human feedback

DialogRPT

Predicting upvotes and replies

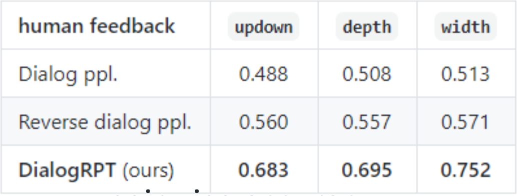
Generic response (e.g. “Me too!”) gets low predicted feedback



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**

Our rankers vs. MMI:



human feedback	updown	depth	width
Dialog ppl.	0.488	0.508	0.513
Reverse dialog ppl.	0.560	0.557	0.571
DialogRPT (ours)	**0.683**	**0.695**	**0.752**

pairwise accuracy

Step 1: 100M + Human feedback data

Step 2: Contrastive learning

Step 3: new dialog rankers!