

Can we detect the need-for-closure bias on social media?

Matthijs Westera & Mark Rademaker



Universiteit
Leiden

'Need-for-closure'?

(Webster & Kruglanski, 1994)

- Cognitive bias similar to, but distinct from, confirmation bias.
- We are more likely to accept incoming information...
...if it answers a question we already entertained.

Research question

When a user encounters a certain piece of information...

henceforth: "the pivot"

...are they more likely to repeat it, if it answers a question to which they were exposed earlier?

Data

- Via Reddit API:
 - r/progressive, r/conservative, r/conspiracy, ...
 - Obtain users from most recent posts.
 - Get those users' posts (with their parent/replies).
- Sentencize all posts/parents/replies. Terminology:
 - **Pivot**: a sentence to which the user responded.
 - **Question**: a question to which the user responded.
 - **Sentence**: any sentence posted by the user
- Heuristically reduced to 'promising' sentences

Currently 580K

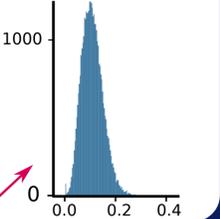
Currently 2M

Currently: 20K pivots, 4K questions, 25K sentences

Challenges

- API rates/limits.
- Sparsity: virtually all posts are unrelated.
- Language style distinct from standard NLI/SQUAD.
- Context: false positives due to missing context.
- No data on users' 'likes'.

histogram of similarity pivot ~ sentence

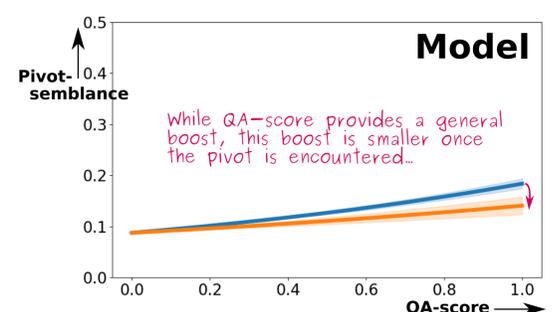


Regression model

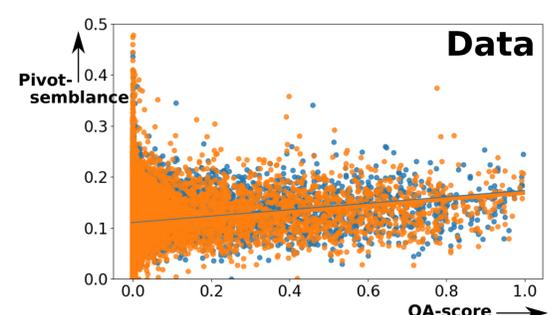
- Variables for each sentence & pivot:
 - **after-pivot**: whether the sentence came after the pivot.
 - **pivot-semblance**: degree to which sentence resembles the pivot. (paraphrase-MiniLM-L6-v2 embedding cosine)
 - **QA-score**: degree to which pivot answers a question. (heuristics + ALBERT) (Alternative: Flan-T5-large, but it struggled...)
- Poisson regression (pymc):
pivot-semblance ~ QA-score × after-pivot
(+ random effects per user & per pivot)

So... No?

- = sentences occurring before exposure to pivot
- = sentences occurring after exposure to pivot



While QA-score provides a general boost, this boost is smaller once the pivot is encountered...



	mean	sd	hdi_3%	hdi_97%
intercept	-2.44	0.01	-2.45	-2.42
fixed_effect	0.74	0.05	0.64	0.84
pivot_effect	-0.28	0.13	-0.52	-0.02