


Analyzing Discourse via Reddit

Seth Ziemke



01. Background

02. Topic overview

03. Motivating
research

04. Project overview

05. Analysis

06. What is next?

table of contents





What is Reddit?



“Reddit is a network of communities where people can dive into their interests, hobbies, and passions”



↑

13.1k


↓


 r/technology · Posted by u/Tough_Gadfly 2 days ago 


Reddit will begin charging for access to its API


techcrunch.com/2023/0...


Social Media




 2.0k Comments

 Award

 Share

 Save

...


5 people here 


Comment as [Primary-Ad3149](#)


What are your thoughts?

B


i










A










...





Markdown Mode

Comment

Sort By: Best ▾

 Search comments

[View discussions in 6 other communities](#)


 ganner · 2 days ago    2


If 3rd party apps stop working that's a great way for me to regain time I waste on this site. Lord knows I use twitter far less than I used to.

↑


10.7k

↓

 Reply

 Share

...


 PERSONA916 · 2 days ago


Have they tried making an app that doesn't suck donkey balls as a way to cut down on API usage? I would use the official reddit app if it was as good as Boost or Relay

↑

754

↓

 Reply

 Share

...

Subreddit

‘Community’ containing numerous posts

Post

Contains information relevant to
subreddit

Comment

Responds to post



Discourse Analysis

Discourse analysis is an extremely broad topic!

- Analysis of language 'beyond the sentence'
- Can take into account the larger context

Multiple approaches for analysis

- **Semantic Similarity**
- **Speech Acts**
- **Frames**
- Turn-Taking



Semantic Similarity

- Given two passages of text, how similar are they in meaning?

“I poured water from the *bottle* into the **cup** until **it** was **full**.”
it => cup

“I poured water from the **bottle** into the cup until **it** was **empty**.”
it=> bottle

[Transformers in NLP: A beginner friendly explanation | Towards Data Science](#)

- Current state-of-the-art models use transformers
- Best available model is ROBERTA-large
 - Trained on 160GB of uncompressed text



Speech Acts

- What does an utterance accomplish?

Speech Act Type	Direction of fit	S=speaker; X=situation
Representative	Make words fit the world	S believes X
Directive	Make the world fit world	S wants X
Commissive	Make the world fit words	S intends X
Expressive	Make words fit the world	S feels X
Declaration	Words change the world	S causes X



Frames

- What is a speaker doing while speaking?
- How do environmental factors contextualize an utterance?
- Framing is important in understanding intent





How do we apply this to Reddit?

With the comments found under Reddit posts, we can:

- View comments in relation to one another
- Look at properties of high-traffic posts



PRAW

Python Reddit API Wrapper

- Can scrape large number of posts with ease
- Data is fresh
- Allows us to build our own dataset





r/AskReddit

id	author	title	text	numComments	comments
122xgeb	kamal31121984	Anyone who works 40 hours a week for minimum w...		25	(jds8enf, jds84e4, jds7zjy, jds936a, jds813w, ...
122yg9v	Routine_Camp7308	What do you consider a successful day for your...		4	(jdscos2, jdsdd8t, jdsd0bz, jdse19r)
122xxz4	BrudderO	What's your favourite way to be seduced?		9	(jdsbg98, jdsa86a, jdsai9g, jdscykh, jdsaadi, ...
122xti7	vicelordjohn	If you were sentenced to death and could choos...		44	(jds9qka, jds9qk3, jds9p0d, jds9pgt, jds9lma, ...
122y4tz	MrFantastic74	How can North America get rid of tipping cultu...		30	(jdsb0k6, jdscykm, jdsb79k, jdsdq3w, jdsb099, ...
122xerf	Impostata	A-Arons of Reddit, how have you done messed up?		3	(jds7snq, jds7xb1, jds7xg0)
122x0ev	FairtalkOfficial	What year do you believe AI will completely re...		26	(jds67us, jds6svt, jds9akd, jds5s28, jds77hg, ...
122xk5t	Agitated-Pilot-1897	What's the best place in your house for hiding...		29	(jds8lc2, jds8gx6, jds8eim, jds8w3o, jdsaqne, ...
122yaam	puce_3000	What's your favorite "grandmas remedy"?		6	(jdsbxdf, jdsbyqg, jdsc2w, jdscf2h, jdscr51, ...
122y2yg	QuietlySmirking	What movies were WAY ahead of their time?		8	(jdsau52, jdsaqmm, jdsasxp, jdsbipf, jdsc1fl, ...



Sample from r/AskReddit

jds67us t3_122x0ev Saw a Star Wars movie years ago that felt like I was watching a cartoon the entire time

jds7os1 t1_jds67us Looks like the AI takeover has already begun if they're making Jar Jar Binks look realistic.

jds6svt t3_122x0ev CGI hasn't put the movie camera manufacturers out of a business. Nor has it made actors obsolete.

AI is going to have a hard time replacing entire industries just like every other technological advancement has. It could change cinema substantially, but it will never kill the camera.

jds9akd t3_122x0ev I don't think it will completely replace it, but it will lower some costs, eliminate some jobs and reduce the time (and probably the quality) of movies in 10-20 years.

jds5s28 t3_122x0ev 2050

jds77hg t3_122x0ev Everyone here, do yourself a favor and watch the "Netflix Bots write movies" series on YouTube by Netflix is a Joke

jds814j t3_122x0ev There hasn't been any type of intelligence, artificial or otherwise, in that industry for many decades. I doubt AI is necessary.

jdsac92 t3_122x0ev I don't get it. Why would a steak sauce replace the film industry?

scsss t3_122x0ev filmmaking is a collaborative process that involves many different people with diverse skill sets and perspectives, and it is unlikely that AI can fully replace the human element of collaboration and teamwork in the industry.

That being said, it is possible that AI and other technologies may continue to evolve and play a greater role in the filmmaking process in the future, but it is unlikely to completely replace the industry.

jds5oy7 t3_122x0ev It won't.

jds5q74 t3_122x0ev It won't.

jds5w4j t3_122x0ev I don't think AI can replace creativity and emotions that humans put into filmmaking.



Analysis

Will focus on the three aforementioned topics:

- **Semantic Similarity**
- **Speech Acts**
- **Frames**

We'll quantify the 'popularity' of a post by the number of replies it receives!





Comment Semantic Similarity

- How similar are the top comments to less popular comments?
- Tested with the ROBERTA model
- Analysis example:

MOST POPULAR COMMENT: Every time you get a raise, hide it. Increase your 401(k) contribution, or put it into an IRA, or invest in stock, or just put it aside in savings.

In other words, don't get used to a higher level of consumption. This is especially important when you're young.

HIGH SEMANTIC SIMILARITY: Don't have kids. 0.6705178022384644

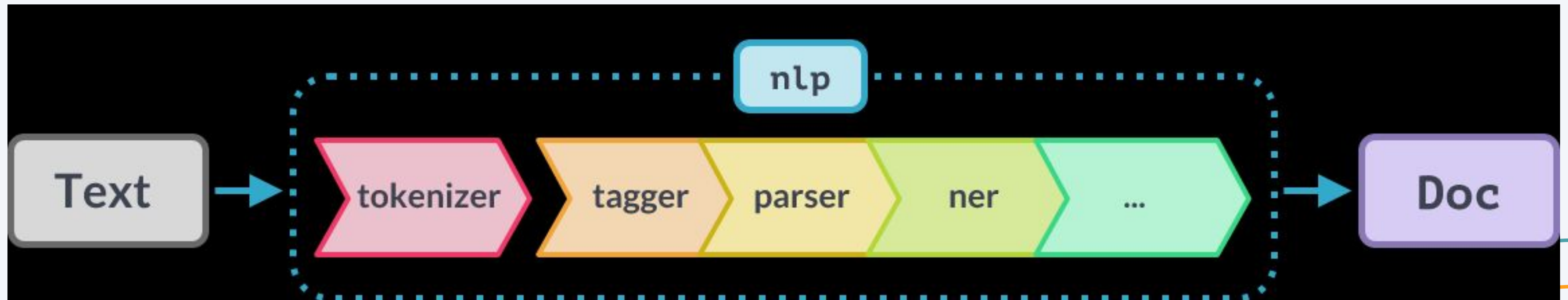
There are a few drawbacks to this model:

- Reddit comments can have varying lengths and levels of detail
- ROBERTA was not trained on internet conversation data



SpaCy

SpaCy is an industry-grade NLP tool with an excellent part-of-speech tagger!





The future...

In the future, I will:

- Further dive into semantic similarity, and see how posts within the same subreddit compare to one another
- Analyze the relationship between token count and popularity of a comment
- Research whether a post containing certain speech acts/patterns results in more engagement



Thank you!

Feel free to reach out to
slz19@pitt.edu with any questions

