

CS 6742  
Reminder: All posted

Agenda: Topic: to what extent is there social interaction on review sites?

Research Skill: introduction to data annotation issues  
Pedagogy: get use to talking to each other comfortable discussing  
research 2: getting inspiration by looking @ data  
in the past, this has been good to say explicitly.

show that refs have been posted.  
- issue: start late b/c of room change?

Warm-up: 1st, tell everyone your name, what langs you speak  
Then, go 'round again, you'll have to pick someone else, say their name, what langs they speak

- come on NLP: soc. interaction => need to go when people are exhibiting lang. together
- Review sites are major source of user-generated-content [UGC]

ex: Amazon (not doing TripAdvisor, [Craigslist; deSist], Yelp, etc.)  
- already on screen - straight Man - excellent academic satire; lots of fun

- people dumping lots of lang there (380 reviews)
- \* always a good idea to browse around, look for research ideas.  
when find out about a new site

- for one thing, observatin that's generated lots of subsequent research:  
stars as useful summary of individual opinion,  
"star statistics" as summary of collective judgement  
[avg, histogram]

researcher's one

⊗ when you see a useful tag, immediate reflex is to think about automating it (hence see if you can predict it)  
The research sub-area of sentiment analysis started in large part from the existence of these stars (thousands of papers on it sense).  
can immediately apply standard supervised ML techniques.

Quick aside: altho sometimes it's really easy: [review entitled "Five stars" print highlighted]  
Sometimes the star annotation seems to diverge from the text

- foreshadowing: user-generated labels may not always be "correct"

"automatically chosen by our system b/c they are representative of what people are saying about them"

- can also ask how to predict these summary quotes [blue highlight]  
(Are they redundant? How are they actually most representative? / Not from most helpful)



→ (e.g. another q: why display reviewer location?  
does "real name" mean pressure to be (+)?

That's all great and interesting, but ~~what about~~ is there any social interaction going on? ( 'Cuz if not, not relevant to this class).

~~Recall from last time:~~

Recall ~~from~~ from last time: two types of settings where interaction via lang. made manifest:

- (1) conversation
- (2) broadcast: social effect

conversation ~~does not~~ can happen ~ yellow highlight: ~~new window~~  
open in new window  
people can comment on reviews  
and people can rply.

Don't know that there's lots of engaged conversation here.  
(sparse - not a good dataset for that)

not for class, but note the hiding of 'non-const.' comment comments.

what about an effect on the others?  
Convenient to split the notion of others:

- (2a) ~~on~~ on other ~~authors~~ reviewers (authors)
- (2b) on non-author readers.

ex: changes what they choose to write about, or whether they choose to write @ all. ('nothing new to add?')

Let's talk about ~~2a & b~~ briefly:

(2a): how could you tell if there is an effect such an effect?

Implications: are you getting "real" opinions?

1st q: do they even read other reviews?

ask them: ~~Gilbert~~

~~At reading they don't!~~ [Gilbert; Karahalios '10]

And, "professional" reviewers don't ~~like the idea~~ thought that reading other reviews is ~~deterministic~~ being asked if they read other reviews is offensive! (altho' phrasing of q played a role)

[Pinch; Kessler '11]: 42% of top reviewers look before  
53% look after  
(5% never look)

look for correlations w/ other reviews.

[Wu; Huberman '10] - later opinions more likely to disagree (effect bc cost of expression is "high", cf. Gilbert; K.)

[Michal; Otterbacher '14] - "herding" effects in the lang. - (adaptation of stg)

[Daneš & Nícolson-Mizl '13] - lang adpt. to group (bear reviews, w/ Amazon)

des. summary what the effect?

in discussion, people came up w/ a lot of these, which is great!

(made pts: can get (+) evidence but not (-) evidence. Also you don't know what order the readers saw the reviews in.

so note: just asking this 1st q has led to a bunch of research, and a best paper award  
but let's now turn to the bigger audience

(2b): need to look for a "reaction"  
in Amazon, comments too sparse.

but, helpfulness annotations as a reaction

< distribute "what constitutes a helpful review" survey? >

< take and shuffle and redistribute

whose is yours; what langs? - reinforce knowing each other >

- show the 42/42 helpful review.

Helpful according to their features?

helpful tally:

not-helpful tally:

{ mostly learned towards helpful.  
but more uncertainty than  
"42/42" would suggest.

Goals:

Purpose: annotation/labeling often an important part of NLP+SI research

But variat'n in interpretation is common

Write instructions to try to expose/control variation

Test-run your annotation instr. first

- did everyone follow your directions?

{ the class ourselves  
was an example of

~~not~~ such a pilot run.

[Then try on undergrads/  
general public, for  
"regular people".

{ often aren't  
lucky enough  
to get  
explicit labels  
AMTurk /  
Crowdfunder,  
etc.

Also, explicit  
labels may  
not mean  
what you  
think.

... we long to see ~~with~~ factors outside lang that affect evals.