

chalk.
dongle
browser tabs
(slashdot viz needs firefox)
printer
Rearrange the chairs; tables

- make sure display ok
- ethernet cable? (wireless connection?)
- be careful about open windows
(non-PhD auditors?)

happy honeyg
last so time pg
A1
Notabilia
Slashdot
URC
visualizer (SD: 1972214)
8/26/09
Lec. 1

CS 6742: <http://www.cs.cornell.edu/courses/cs6742>
Assignment 1 (A1) has been posted

Today's course overview

~~Non-PhD students~~

Hi, and welcome to the 1st meeting of CS6742, "Natural language processing and social interaction".

This is a subject ~~that~~ that's ~~generating~~ generating a lot of research interest these days, ~~especially with the rise of social media~~, and I'd first like to say a few words about why that is.

⊙

To start with, we all know that (spoken; written) language is one of the main channels by which we human beings communicate with each other.

• That ~~statement~~ ^{assertion} is probably so well-accepted that some of you might be thinking ~~that~~ right now that language is the only such channel, but that's not true:

There's facial expressions
posture
gesture
actions, and even inactions, if you think about it

~~But language~~

But still, language is clearly one of the key ~~ways~~ ^{channels} by which we

- make group decisions
- maintain relationships
- and, to be more egotistical about it, accomplish our own goals when other people are involved.

Now that's always been true throughout human history.

~~What's changed now is that we have tons of digital records of language interactions~~

~~tons more of those language~~
tons of language data is now available ~~to us~~ computationally accessible to us.

it's diff. about now?

key thing is:

the development of sites and systems that make language-based interaction manifest.

purpose: point: show interesting examples as context in which to make points.

Need outline on board because will go deep in 'conversation' ~~tree~~ discussion tree [with any web]

I mean more by that than what might be obvious. What you probably first thought of when you heard or read a phrase like that would be what one might call "~~conversations~~"

"conversation" ~~systems~~ systems/sites (s/s)

~~"private audience"~~: texts, IMs, emails, ~~recorded conv~~

- analogs of traditional forms

- ~~interaction-inclusive~~ ~~eg. FB, Twitter~~

(of meetings) here I don't mean in the surveillance sense, altho' those exist too, but in the meeting ~~to see~~ transcripts
[one spectacular version relatively recently]

"social media"

goal = interaction (FB, Twitter, Reddit, youtube comments)

goal = task (~~butipedia talk pages~~) ~~critic~~ ~~talk~~ discussion pages on wikis

(note wks own stat that it is not a social networking service)

~~"broadcast and effect" s/s~~

~~est~~ arguably but another, less obvious type of development that's made MLP: social interaction ~ a hot topic is:

"broadcast" + social effect ~~s/s~~ ~~data~~ ~~recombination of s/s~~

And note that I'm contrasting these w/ conversations:
ex: ~~how does~~ ~~how does~~ framing of an issue can affect attitudes; actions

tex: effect on voter turn-out of the phrasing of the appeal

[Bryan, Rogers, Walton, Rogers, Dweck PNAS 2011]

- ~~to study this~~ • appeal - attempting to "get out the vote"

- two variants of the appeal:
either stress: "be a voter" (identity)

action: go out and vote

- assessed by, among other things, checking the public records of whether the people actually voted.

Here, it's not a single s/s involved, but I hope you get the idea: we have access to data that can help us assess the social

effects of language and I'm hinting here that language ^{can} matters a lot in soc. int.

of, so you can see that we're not going to talk just about conversations in this course.

But that being said, let me now show you two examples of ^{interesting} conversational sites, [really, visualizations of them], to give you an idea of what kind of possibilities there are.

~~Two examples of conversational sites~~

Two examples of conversational ^{issues} ~~explorations~~.

- (a) voting discussion (Wikipedia) = (good is task)
- (b) social discussion (moderation, structure; Slashdot) = (good is interaction) } back to hierarchical type

NOTE: lots of implicit annotation → a research "jumpstart"!

Notable ~~met~~:
Wikipedia has lots of behind-the-scenes work & discussion by editors

skin Wc will discuss later;
but intro Slashdot b/c of reading
start w/ Wk b/c we just talked about voting

→ (a) "voting" on whether Wk should keep an article on interchanges in a particular highway.

→ ~~the argument~~ so, consequential discussion (altho' Wk votes are not necessary)

- in fact, at least for some Wk "votes" it's actually not necessarily a 'majority' vote but to convince another authority.

"consensus is not based on a tally... but on reasonable, policy-based arguments"

highlighting via Diigo

informing;

"interceptions": / That "roadcraft" must exist as a term: / ^{lexical innovations are an interesting cue!}

- note: replies (sometimes labeled as comments) - ① blue highlights
- note: back-"edit" - Happy Camper, Jaranda, Happy Camper adds. ② pink highlights
- : re-entry of "Factious" - a back-n-forth. others are "guestbook-y". ③
- : change of opinion given strikethrough and, change of opinion based on argument.
- ③ green highlights

some are snarky ("Gag").

→ notabilia.

other observations... blue highlights

* App. pt of view: produce ~~is~~ summary of how discussion is going, main points?

[not for now: the inclusion of username: time stamp is optional, I think]

* Analysis: I think you can ~~ask~~ ^{study} some pretty interesting ~~questions~~ ^{things} here, like whether there's stuff in the laws that influences the dynamics

(runs of votes; sudden 'popularity' change in votes)

→ ~~notabilia~~ actually a run of one side. / get branch from left

- and learning what kind of arguments are most convincing is useful for real life, btw.

→ now let's go to our (b) example, a social conversation on slashdot.

- ~~less purposeful~~, - more structured; explicit parenting of comments
- not "consequential", attracts "too" many comments [at bottom: Get 109]
~~some perhaps less valuable~~
- system tries to let you see the 'best' comments, or @ least select them.

- scores: - show range selector

- tags: so can skip the 'funny' stuff.

• ~~voting by populace/moderators~~
applied by moderators (complicated)
(complicated: users become moderators)

* From an app point of view

- producing ~~an~~ summary, giving an overview of the conversation.

→ **UBC**: using NLP to show sentiment (neg or pos), topics. threads by indent
beyond comment ~~length~~ thread length. ^{inc individual keywords:}
- could we see that a discussion is becoming non-useful? } automatically
or is a good one that people should be aware of.
→ @ post length + sentiment distribution might show the "most thoughtful" ones?

→ **Kaltenbrunner**:

value in comments that sport "trees": - can we predict? Fat vs. skinny trees?

There are also interesting questions from a social psych point of view, too.
What cuts off a thread?

Now, given that we have all the interactions manifested in natural-lang. format, what can we do with them?

~~The~~ We'll be pursuing two opportunities in this course:

Two opportunities of ~~interest~~:

1) Analysis: use NLP to understand interactions (and interactors) better
ex: are there language cues that correlate with successful interactions?

(non-) ... ideas?
↑
(phrasings of)

can we learn which participants are engaged / ^{crucial} ~~in power~~ / agree?
(dis) (not) (dis)

~~usual~~ standard approach: mining from a large-scale ^{corpora} ~~corpus~~
(“bodies” of “text”)

2) Facilitation: use NLP to improve interactions, or cause them

exa: make a hidden state evident (confusion, antagonism, etc...)
Maybe even intervene?

~~I want to bring to your attention a particular w.~~

It's nice when these go hand-in-hand.
2 opps.

~~But speaking of~~

But speaking of opportunities, I want to take this opportunity to talk about another facet of this course.

⊙ This word “opportunities” here — notice it doesn't say “solutions”.

This is meant to indicate that ⊙ this class is a

↓ research-oriented course

⊙

Just did context -goals ; some examples.

~~Cont~~

↓ About pedagogical goals

~~And the reason is that if that's fair to say that then~~

~~That's~~ The reason is that we haven't made
This ^{class} can't be a survey of ~~approx~~ known approaches to this problem
b/c there's ~~no~~ ~~canonical~~ aren't really a set of established approaches
yes.

... just understanding processing language "in the wild" is hard.
no "canonical answer".

③ Example review excerpt:

"Read the book" → positive or negative opinion?
< J: negative sentiment toward interlocutor >

[Laugh due to Bob Black] ~5 people didn't see, all non-native speakers of English.

② Or, "She runs the gamut of emotions from A to B" →
① Or, "If you like this fragrance, please wear it @ home] and top your windows shut"

so be proud of yourself for being able to understand language @ all.

→ meaning is clear to you.
→ you have to know that scents & smells can be triggered this way.

→ you have to know that the word phrase is part - for A to B.

→ even a human can't tell b/c of context dependence.

OK, so that's one reason this is a research-focused course, is because the very topic matter is one in which, in a sense, very little is known.

There's also a second reason why this is a research-oriented course, and ~~that's~~ it's b/c I know not all of you are going to do research in NLP for a living,

~~but~~ ~~and~~ 6000-level courses are for all PhD students, when PhD students are people who are planning on doing research for a living -

why research #2: ... 6000-level courses are training for PhD-level research. - research methodology.

So much of the way this course is organized is driven by the desire to provide training in the research process, whatever your eventual research interests turn out to be.

So the end product will be a research project, and the goal is throughout the semester is to get work through that process of doing a research project.

How does this agenda inform the course structure?

First, the early part of the course is going to interleave presentation of fundamental material with you doing a pilot research study right off the bat.

- show webpage. show prior year to demonstrate lots of info
- show course structure

g's about what the fundamentals will be
NOT NLP fundamentals
(parsing, WSD)

~~will be doing~~

So as you can see, your first asst is released as of today.

~~will be doing~~ AI

~~you're tasked w/ coming up with~~

~~It asks you to engage in a pilot~~ come up with

It involves 2 readings, one on reviews and one on comment threads.

~~you're~~ You're tasked w/ coming up with a research idea based on those readings, and do a pilot empirical study.

Notice that ~~at the~~ simultaneous with that assignment,

~~will be~~
I'll

be doing some lectures on online reviews and conversation threads, together w/ some gray discussion of the readings and/or how your pilot studies are going.

After that...

- some bkgd lectures
- then some student-led versions of AI, where we read papers; try at some research ideas inspired by them
- and then there's the final-project portion of the course.

→ Eugene Charniak: 'do sthg'

→ papers chosen by me to be accessible, thought-provoking.

COMING UP WITH PROBLEMS

AI is representative of the spirit of the course, so included or plan

→ AI.

ended @ about 11:10.
About 30 showed up.

~~Non-PhD~~
Masters, grad, a doctors

Piazza = lots of feedback! suggestions - don't be worried that you don't know anything.

Confidence-building of the MIT area exam.