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

Today's course overview

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

This is a subject that's generating a lot of research interest these days, especially with the rise of social media, and I'd 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 is probably so well-accepted that some of you might be thinking right now that language is the only such channel, but that's not true.

Then's facial expressions, posture, gestures, actions, and even inactions, if you think about it.

But, language.

But still, language is clearly one of the key channels by which we:
- make group decisions
- maintain relationships
- and, to be more specific, about it, accomplish our own goals when other people are involved.

Now that's always been true throughout human history.

What's changed, more is that we take two of digital records of language interactions.

Thousands of language data is now available computationally accessible.
It's different about now?
The key thing is:
The development of sites and systems that make
text-based interaction manifest.

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

"conversation" systems/sites (s/s)
- "private audience": texts, IMs, emails
- analogs of traditional forms
- interaction inclusive (e.g., FB, Twitter)
- transcripts

"Social media" goal = interaction (FB, Twitter, Reddit, yahoo!_groups)
goal = task (Wikipedia talk pages) initial/Fb discussion pages on wikis

One spectacular version relatively recently

"Broadcast + Social effect": s/s
- arguably
- but another, less obvious type of development that's made MP: social
interaction a hot topic is:

"Broadcast + Social effect": data

And note that I'm contrasting these with conversations:
Framing of an issue can affect attitudes + actions
- ex: how does framing of text:
- ex: effect on voter turnout of the phrasing of the appeal
[Byrne, Rogers, Walton, Rogers, Duwe, PNAS 2011]

- attempting to "get out the vote" - 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 simple 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 with a lot in soc. in.
Two examples of conversation sites

Two examples of conversational explanation:
(a) voting discussion (Wikipedia) — consequential, go/no go
(b) social discussion (moderation, structure, Slashdot) — (goal is interaction)

Shim: [NOTE: lots of implicit annotation] a research "jumpstart!"

We will discuss later but intro Slashdot b/c of reading

(a) Voting on whether we should keep an article in the exchanges in a particular highway.

(b) Anecdotes so, consequential discussion (attributed votes are not necessary)

- In fact, or at least for some types, "votes" is actually not necessarily a "majority" vote but to convince another authority

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

"Interceptions": That "roadcraft" must exist as a term. Some lexical innovations are on interesting cue!

Note: replies (sometimes labeled as comments)


- Re-entry of "Fachtwiss": a back-and-forth, others are "guestbook-

- Change of opinion given strikeout and change of opinion based on argument.

Green highlights

Other observations: Blue highlights

* App. pt of view: produce summary of how discussion is going, main points? study of things (run of votes suddenly change in votes)

* Analysis: Is there nothing here like what’s shuff in thelangs that influences the dynamics

[Note for now: the inclusion of us names is optional, I think].
- 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

- system needs to let you see the 'best' comments, or at least select them.

- scores: show range selector
- voting by popbase/moderators
- tags: so can skip the 'vomiting' stuff, applied by moderators (complicated: users become moderators)

From an app point of view: producing one.

Similarly, giving an overview of the conversation.

Use: using NLP to show sentiment (neg or pos) - topics, threads by indent length, could we see that a discussion is becoming un-useful? Automatically.

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

Then an also interesting question 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?

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 / degree?
   (dis) (not) (dis)

   used standard approach: mining from large-scale corp.
   ("booster of text")

2) Facilitation: use NLP to improve interactions, or cause them
   ex: make a hidden state evident (confusion, antagonism, etc...)

   Maybe even intervene?

I want to bring to your attention a particular one:
It's nice when these go hand-in-hand.

But speaking of opps.

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 contest -goals; some examples:

About pedagogical goals
And the reason is that it's hard to say that
there's no "canonical answer".

This isn't a survey of approaches to this problem;
there isn't really a set of established approaches yet.

... just understanding processing language "in the wild" is hard.

Example review excerpt:

"Read the book" positive or negative opinion?

[John's negative sentiment toward excerpt]

[Review due to Bob] J didn't see.

Or, "She was the gamut of emotions from A to B.

Or, "If you like this fragrance, please wear it @ home.

[You have to know that scents & smells can be sprayed into]

[You have to know that the word "peppermint" is used for A to B.

Even if you can't tell a bit of context detail.

OK, so that's one reason this is a research-focused course.

It's because the very topic matter is one in which, in a sense, very little is

Ther's also a second reason why this is a research-oriented course.

And that's because I know not all of you are going to do research in NLP for a living.

Both 6000-level courses are for all PhD students,

whom PhD students are people who are planning on doing research for a

why course #2:

... 6000-level courses are training for PhD-level research.

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 thought to

form a useful tool for the process of doing research.
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 webpg; show your year to demonstrate lots of info
- show course structure

*1st talk about subject matter
*qs about what the fundamentals will be

NOT MLP fundamentals
Not WSD

Will be doing

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

HEALING AI

you'll need to continue with
It asks you to engage with come up with
It involves 2 readings, one on reviews and one on comment threads.
You're tasked w/ coming up with a research idea based on those readings, and do a pilot empirical study.

Notice that simultaneously w/ that assignment,
I'll be doing some lectures on online reviews and conversation threads,

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

AI is representative of the spirit of the course, so

Piagge: lots of feedback
Suggestion: don't be worried that you don't know anything

Confidence building of the MIT area exam.

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