- remote
- ~~chalk~~
- wb markers
- display dongle
- <is there a clock??>
- black screen

← Piazza: topic/paper suggestions

Lecture 1

8/29/13

CS 6742 Assignment 1 (A1) has been posted : http://www.cs.cornell.edu/courses/cs6742
Colloquium (research) talk by me today, 4:15 B17 Upson
~~Auditors; non-PhD students: see me after class~~
(just say TI

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

I'm the instructor; I'm a professor in computer science and in information science, and these days,

~~NLP and social interac~~
research in NLP and social interaction is what I do for a living,
(and for fun, come to think of it).

I'd give you more specifics now, but
(point)

I'm ~~po~~ giving the CS department colloquium today,
where I'll be talking about two recent projects I've been involved
in that I'm really excited about,
so please come to that talk if you'd like to hear about what I've
been up to lately.

~~Assistant~~
As for this course and why it exists:

We all know that spoken & written language is one of the main
channels by which we human beings communicate with each other.
(If you're wondering what some of the other channels are,
there's facial expression
gesture
actions ~~or~~ and ~~p react~~ even inactions, if you think about it
posture ...)

Anyway, about all this language-based communication: e
I doubt it's escaped your notice that more & more of such
communication is becoming computationally accessible, since it's
online.

Just to reel off ~~some~~ one factoid(s):

As of June 2012, Twitter was seeing 400 million tweets per day ~~(CEO of Twitter~~
@ [CEO of Twitter @ an Economist event]

(1 m, 10 m, 100 m)

~~In 2011, typical corporate email user sends & receives ~105 emails/day.
About 19% getting to mailbox are ~~spam~~ + graymail (unwanted notifications/newsletters).
(This is actually lower than before b/c of IMs/social networks)~~
[Radicati report]

We can ~~outline some possibilities~~ perform a rough categorization of the kinds ~~that this computational accessibility of conversations implies~~ of conversational interactions that are computationally accessible; ~~although~~ I don't mean the following ~~to~~ classification to necessarily be taken too seriously; but ~~it do~~ taking this perspective allows one to gradually realize how broad the playing field is. ~~when we first think of conversation, what we spring to mind is:~~

"private audience" (analog of f2f conversations, @ ~~@~~ 1st blush):
small→ IM's, chats, emails.
or @ least intended.

writing down so to indicate that students should take notes.

~~Dis class~~ The classification I'm going to present is based on audience b/c intended audience is an important ~~out~~ factor in communication. Even for diaries that are meant to be totally private and for no one else's eyes — what do diary entries often start with?

↓ people thought of diaries, private blogs. But also liked the joint about "dear day."
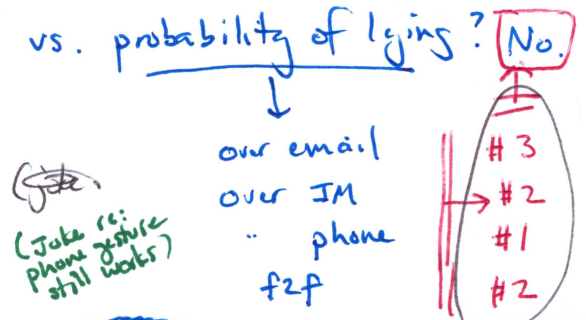
Now, actually, the differences and similarities wrt f2f conversations is an interesting topic that we won't get into today.

But just to point out some interesting work done @ Cornell on the subject,

Let me briefly mention some work done by Jeff Hancock and colleagues;

Jeff is a professor of communications.

including a lie

Aside: "distancing" of medium vs. probability of lying? [No.]

(Joke re: phone gesture still works)

over email   #3
over IM   #2
" phone   #1
f2f   #2

but nature of lie must be taken into acct: "gotta go"

[Hancock, thom-Santelli, Ritchie '04]

- ~~I'll post lecto.~~
lecture refs → course homepg

• OK, so but anyway, ~~there are these~~ yes, many of the conversations we have are one-on-one. ❶ in the online or offline work

But ~~your pretty~~ we also have group conversations pretty frequently, where more people are involved, and that may also be semi-public depending on when they take place

private/public audience.
~~private~~ • ~~private-ish intended audience, public location~~ ❷ = group discussions:
  where do these take place online?
  social-network conversation threads
  blog comment threads

Q/A ~~...~~ → when "blog" here is broadly construed
  some may be meant for ~~...~~ subset of participants ~~...~~
  ~~quotation marks ... public ...~~
  but may also be an element of the public being an
                    intended audience, as well.

~~OK, so if we've written~~
And that thought leads us to:
  public audience ≈ • analog to 'soap boxes'
    blog posts (even w/out comments)
    what else?
    one you might not think of:
                              reviews on
    review aggregation sites (Yelp, Amazon, etc.)
      • now, here, you may feel like challenging me:
  ads   That is, you might ask, "~~how~~ does someone leaving a
discussion sites   review on Amazon ~~...~~ really qualify as a kind of
                    social interaction?"

      ~~And it~~
                                        argue that we should
      Not surprisingly, I'd like to ~~say yes~~ consider that
      within scope.
      The more obvious way to see this is to ~~can~~ think of
        such a review as • an act that may cause so. else to
                      – adopt an opinion
                      – change their mind
                      – make a decision, such as to buy.
      But you'll see another perspective in the reading I'll
        be assigning to you today / More on that later.
                                      ↑ in future lectures and

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)ᶺ   (not)ᶺ   (dis)ᶺ

~~usual~~
standard approach: mining from large-scale corpus
                                  a   us
                          ("bodyies" 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 w:~~
It's nice when these go hand-in-hand.
                        ↳ 2 opps.

~~But speaking of op~~

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

∅
         :

And ~~the reason is that I that it's fair to say that there's~~

~~That's The reason is that to have made~~

This ~~class~~ ~~can't~~ can't be a survey of ~~approx~~ known approaches to this problem b/c there's ~~so~~ ~~can't~~ ~~so~~ aren't really a set of established approaches yet.

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

⑧ Example review excerpt:

"Read the book" → positive or negative opinion?

[example due to Bob Bland]

~) ~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 tape your windows shut".

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

you have to know that the normal phrase is gamut -- from A to Z.

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

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

O.K. 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~~ ~~but~~ 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.

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 good research~~ ~~choosing~~ research projects

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 prior year to demonstrate lots of info
- show course structure

1st talk about subject matter

~~Will be doing~~

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

~~It asks you to engage, inspect come up with~~

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

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

Notice that ~~do that~~ simultaneos with that assignment,

~~Will be~~ ~~will~~ I'll be doing some lectures on online reviews and conversation threads, together w/ some group discussion of the readings and/or ⊙ how you pilot studies are going.

After that...
- some bkgd lectures.   ↑ Eugene Charniak: 'do sthg'
- 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.

AI is representative of the spirit of the course, s.
                  intended or planned

‖→ AI.

~~End~~

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

~~PhD~~, Masters, ugrad auditers ?!