A Tempest

Or, on the flood of interest in:
sentiment analysis,
opinion mining,
and the computational treatment of subjective language

Lillian Lee
Cornell University
http://www.cs.cornell.edu/home/llee

"Romance should never begin with sentiment. It should begin with science and end with a settlement."

— Oscar Wilde, *An Ideal Husband*

O brave new world, that has such people in't

People search for and are affected by online opinions.

TripAdvisor, Rotten Tomatoes, Yelp ...; Amazon, eBay, YouTube...; blogs, Q&A and discussion sites, ...

58% of US residents have researched products online.

73%-87% of review readers say that the reviews were significant influences.

Consumers report being willing to pay 20%-99% more for a 5-star-rated item than a 4-star-rated item. [Comscore '07, Pew '08]

O brave new world, that has such people in't

People search for and are affected by online opinions.
TripAdvisor, Rotten Tomatoes, Yelp ...; Amazon, eBay, YouTube...; blogs, Q&A and discussion sites, ...

58% of US residents have researched products online.

73%-87% of review readers say that the reviews were significant influences.

Consumers report being willing to pay 20%-99% more for a 5-star-rated item than a 4-star-rated item. [Comscore '07, Pew '08]

But 58% of US internet users report that online information was missing, impossible to find, confusing, and/or overwhelming.

Creating technologies that find and analyze reviews would answer a tremendous information need.

Other business applications

As part of brand management, business intelligence systems could:

- search out, analyze, and summarize opinionated mentions of a company's products
- automatically process customer feedback

Beyond consumption: politics

In 2008, 44% of US residents (60% of Internet users) used the Internet to get news and information about politics or the election [Pew '09].

- 33% said that *most* sites they use *share* their point of view.
  21% said that most *challenge* their point of view.
  (See `http://www.slideshare.net/PewInternet/social-media-week-feb-2013-v2` for updates)

Beyond consumption: politics

In 2008, 44% of US residents (60% of Internet users) used the Internet to get news and information about politics or the election [Pew '09].

- 33% said that *most* sites they use *share* their point of view. 21% said that most *challenge* their point of view. (See `http://www.slideshare.net/PewInternet/social-media-week-feb-2013-v2` for updates)

Governmental eRulemaking initiatives (e.g., www.regulations.gov) directly solicit citizen comments on potential new rules

- 400,000 received for a single rule on labeling organic food

Connections to other fields

- Economics: "Does a positive reputation increase revenue, or is word of mouth enough?"
- Political science: "How has public opinion about the death penalty changed, and what are the causes?"
- Sociology and social psychology: "How are people's opinions influenced by the opinions of their friends?"

(More on connections later.)

A perfect storm

There has been a huge upswell of activity in *sentiment analysis* ...
(a.k.a. opinion mining, subjectivity analysis, review mining, etc.)

... due to the many applications just mentioned,
plus improvements in machine learning methods and the new availability of relevant datasets,
and the inherent interestingness of the research problems involved.

- At least 2700 papers since 2002 [Google scholar]

What's past is prologue; what's to come

Part I Introduction

Part II Challenges: a few examples showing the complexity of the language phenomena involved

Part III Selected algorithmic ideas: domain adaptation, modeling label trajectories, and collective classification

Part II: Challenges

The “easiest” sentiment-analysis problem: Classify an avowedly subjective text as either positive or negative (“thumbs up or “thumbs down”).

One application: summarization.

> Elvis Mitchell, May 12, 2000: *It may be a bit early to make such judgments, but Battlefield Earth may well turn out to be the worst movie of this century.*

Part II: Challenges

The "easiest" sentiment-analysis problem: Classify an avowedly subjective text as either positive or negative ("thumbs up or "thumbs down").

One application: summarization.

> Elvis Mitchell, May 12, 2000: *It may be a bit early to make such judgments, but Battlefield Earth may well turn out to be the worst movie of this century.*

Can't we just look for words like "great", "terrible", "worst"?
Yes, but ...
...learning a sufficient set of such words or phrases is an active challenge. [Hatzivassiloglou and McKeown '97, Turney '02, Wiebe et al. '04, and many more]

Results from a pilot human study [Pang, Lee, & Vaithyanathan '02]

	Proposed word lists	Accuracy
Subj. 1	Positive: dazzling, brilliant, phenomenal excellent, fantastic Negative: suck, terrible, awful, unwatchable, hideous	58%
Subj. 2	Positive: gripping, mesmerizing, riveting spectacular, cool, awesome, thrilling badass, excellent, moving, exciting Negative: bad, cliched, sucks, boring stupid, slow	64%

Results from a pilot human study [Pang, Lee, & Vaithyanathan '02]

	Proposed word lists	Accuracy
Subj. 1	Positive: dazzling, brilliant, phenomenal excellent, fantastic Negative: suck, terrible, awful, unwatchable, hideous	58%
Subj. 2	Positive: gripping, mesmerizing, riveting spectacular, cool, awesome, thrilling badass, excellent, moving, exciting Negative: bad, cliched, sucks, boring stupid, slow	64%
Auto	Positive: love, wonderful, best, great, superb, beautiful, **still** Negative: bad, worst, stupid, waste, boring, **?**, **!**	69%

(Nb: accuracies in the 90s have been achieved on this data)

Beyond indicative terms

- 1. This laptop is a great deal.
  2. The release of this laptop caused a great deal of hoopla.
  3. Yeah, this laptop is a great deal ... and I've got a nice bridge you might be interested in.

Beyond indicative terms

- 1. This laptop is a great deal.
  2. The release of this laptop caused a great deal of hoopla.
  3. Yeah, this laptop is a great deal ... and I've got a nice bridge you might be interested in.

- This film should be brilliant. It sounds like a great plot, the actors are first grade, and the supporting cast is good as well, and Stallone is attempting to deliver a good performance. However, it can't hold up.

Beyond indicative terms

- 1. This laptop is a great deal.
  2. The release of this laptop caused a great deal of hoopla.
  3. Yeah, this laptop is a great deal ... and I've got a nice bridge you might be interested in.
- This film should be brilliant. It sounds like a great plot, the actors are first grade, and the supporting cast is good as well, and Stallone is attempting to deliver a good performance. However, it can't hold up.
- The phrases "amazing camera" and "decent quality" have *negative* effects on demand [Archak, Ghose, and Ipeirotis '07]

Beyond indicative terms (continued)

- She ran the gamut of emotions from A to B. [Dorothy Parker, describing Katharine Hepburn]
- Read the book. [Bob Bland]

Part III: Three algorithmic ideas

So, “just” polarity classification has proven harder than topic classification [Pang, Lee, and Vaithyanathan '02].

We'll now look at three particularly interesting ideas that have been applied to sentiment analysis.

Theme: relationships (btwn. domains, labels, and/or items)

Setting for idea 1

Sentiment features for one domain often don't generalize to another. [Turney '02, Engström '04, Read 05, Aue and Gamon '05, Blitzer, Dredze, and Pereira '07, etc.]

- "Read the book."
- "Unpredictable" (movie plots vs. car's steering) [Turney '02]

Domain adaptation via structural correspondence learning

Blitzer, Dredze & Pereira '07 [cf. B., McDonald & P. '06, Ando & Zhang '05]

Book reviews (source domain: labeled)	Kitchen appliance reviews (target domain: unlabeled)
(–) terrible and predictable	terrible leaking thing
(–) absolutely terrible	truly terrible
(+) great	OK
(–) so predictable it's terrible	terrible, always leaking

Domain adaptation via structural correspondence learning

Blitzer, Dredze & Pereira '07 [cf. B., McDonald & P. '06, Ando & Zhang '05]

Book reviews (source domain: labeled)	**Kitchen appliance reviews (target domain: unlabeled)**
(–) terrible and predictable	terrible leaking thing
(–) absolutely terrible	truly terrible
(+) great	OK
(–) so predictable it's terrible	terrible, always leaking

1. Choose *pivot features* that are relatively frequent in *both* domains (need sufficient data) and have high mutual information with the source labels (want "terrible", not "the")

Domain adaptation via structural correspondence learning

Blitzer, Dredze & Pereira '07 [cf. B., McDonald & P. '06, Ando & Zhang '05]

Book reviews (source domain: labeled)	**Kitchen appliance reviews (target domain: unlabeled)**
(–) terrible and predictable	terrible leaking thing
(–) absolutely terrible	truly terrible
(+) great	OK
(–) so predictable it's terrible	terrible, always leaking

2. Encode correlations between pivot and non-pivot features features to create a new feature projection (via the SVD) *to learn target-domain features that act like source-domain features.*

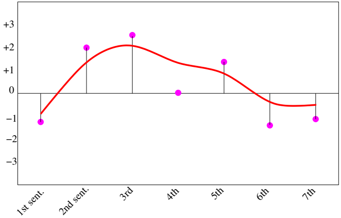
Setting for idea 2

"1 star" is more like "2 stars" than like "5 stars", whereas topics don't typically seem to lie on a continuum.
(There are other interesting structures to sentiment labels, too.)

Approaches include ordinal regression [Wiebe, Wilson, and Hwa '04], classifier stacks [Koppel and Schler '05], metric labeling [Pang and Lee '05], and generalizations of metric labeling [Goldberg and Zhu '06].

Sentiment flow

Mao and Lebanon ['07, '09] combine a sequence of local sentiment judgments into a sentiment flow, and apply nearest-neighbor classification to the *document flows*.



Isotonic conditional random fields impose monotonicity constraints that are consistent with sentence labels' ordinal relationships.

Documents are represented by the "flow" of their discourse.

Collective classification

Relationships between items can be a rich source of information about for performing classification or regression on the items.

- Nearby sentences can share the same subjectivity status, subjective or objective [Pang&Lee ’04]
- Mentions separated by “and” usually have similar sentiment labels; those separated by “but” usually have contrasting labels [Popescu&Etzioni ’05, Snyder&Barzilay ’07]; similar reasoning holds for synonyms and antonyms [Hu&Liu ’04]
- In some domains, references to other speakers generally indicate disagreement [Agrawal et al ’03, Mullen&Malouf ’06, Goldberg, Zhu & Wright ’07] (cf. Adamic&Glance [’05])

Speaking of which ...

A matter of debate

Thomas, Pang, and Lee ['06]:

Given: transcripts of U.S. Congressional floor debates

Goal: classify each *speech segment* (uninterrupted sequence of utterances by a single speaker) as supporting or opposing the proposed legislation

Important characteristics:

1. Discussion context: some speech segments are responses to others
2. Very wide range of topics: flag burning, the U.N., “Recognizing the 30th anniversary of the victory of U.S. winemakers at the 1976 Paris Wine Tasting”
3. Presentation of evidence rather than opinion (see also [Kim&Hovy '06])

   “Our flag is sacred!”: is it pro-ban or contra-ban-revocation?

4. Ground-truth labels can be determined automatically (speaker votes)

Using discussion structure

Two sources of information (details suppressed):

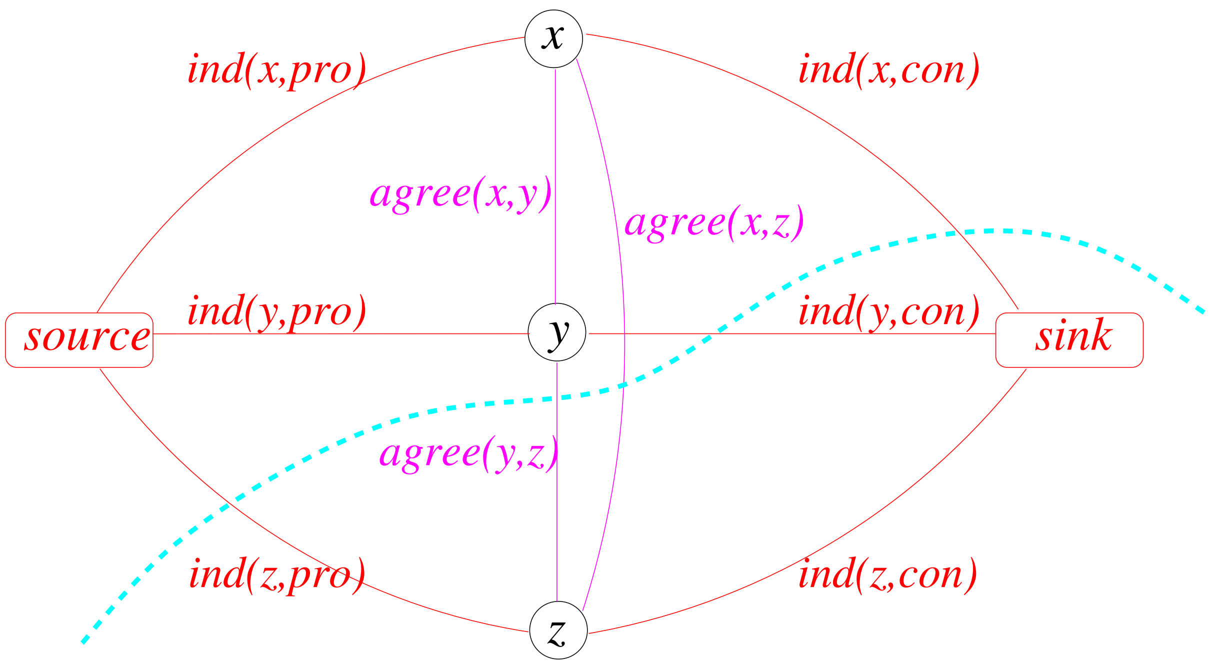
- An individual-document classifier that scores each speech segment x in isolation
- An *agreement classifier* for *pairs* of speech segments, trained to score by-name references (e.g., "I believe Mr. Smith's argument is persuasive") as to how much they indicate agreement

Optimization problem: find a classification c that minimizes:

$$\sum_{x} \textit{ind}(x, \overline{c}(x)) + \sum_{x,x':\, c(x) \neq c(x')} \textit{agree}(x, x')$$

(the items' desire to switch classes due to individual or associational preferences)

A "mitosis" encoding



When the edge weights are non-negative, *network-flow techniques find the min-cost cut efficiently and exactly*.

Negative agreement weights can be handled via pre-processing heuristics [Bansal, Cardie & Lee '08]

Economics: Sample quote (much debate in the literature)

> *If [the estimated] percentages are applied to all of eBay's auctions [$1.6 billion in 2000 4Q], ... sellers' positive reputations added more than $55 million to ... sales, while non-positives reduced sales by about $15 million.*
> *[Houser and Wooders '06]*

Economics: Sample quote (much debate in the literature)

> *If [the estimated] percentages are applied to all of eBay's auctions [$1.6 billion in 2000 4Q], ... sellers' positive reputations added more than $55 million to ... sales, while non-positives reduced sales by about $15 million.* *[Houser and Wooders '06]*

⇒ incentives for manipulation. How can we counteract it?

- Detection of fake reviews [Ott, Choi, Cardie, Hancock '11]
- Algorithms for low-regret product selection from recommendations under adverse conditions [Awerbuch and R. Kleinberg '05]

Read the book

Even more information about applications, research directions, connections to other fields, and other matters is available ...

Opinion Mining and Sentiment Analysis

Bo Pang and Lillian Lee

www.cs.cornell.edu/llee/opinion-mining-sentiment-analysis-survey.html

135-150 pp, 330+ references, in print by next week, we'll also post the pdf when ready

Includes bibliographies, pointers to datasets, more snazzy examples, etc.

Our revels now are ended

We have seen that sentiment analysis...

...has many important applications

...encompasses many interesting research questions

...extends to many areas

You can start working on sentiment analysis right now!

Publicly available datasets include those at http://www.cs.cornell.edu/home/llee/data .

Our revels now are ended

We have see that sentiment analysis...

...has many important applications

...encompasses many interesting research questions

...extends to many areas

You can start working on sentiment analysis right now!

Publicly available datasets include those at http://www.cs.cornell.edu/home/llee/data .

This is such stuff as dreams are made on!