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

People search for and are affected by online opinions.

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

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, ...

According to a Comscore '07 report and an '08 Pew survey:
60% of US residents have done online product research.
15% do so on a typical day.

73%-87% of US readers of online reviews of services say the reviews were significant influences. (more on economics later)

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, ...

According to a Comscore '07 report and an '08 Pew survey:
60% of US residents have done online product research.
15% do so on a typical day.

73%-87% of US readers of online reviews of services say the reviews were significant influences. (more on economics later)

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.

Beyond consumption: politics

In 2006, 31% of US residents used the internet for gathering or sharing political information (60M+ people).

- Major reason?
  28%: to get perspectives from *within* their community.
  34%: to get perspectives from *outside* it.

- 28% said that most sites they use *share* their point of view.
  29% said that most *challenge* their point of view.

[Rainie and Horrigan Pew survey, '07]

Beyond individual interest

Business intelligence systems could ...

- search out, analyze, and summarize opinionated mentions of products, features, consumer desires, etc.
- automatically process customer feedback

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

Many other applications exist, as well.

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 (discussed soon).

- Many companies work on sentiment analysis systems (e.g., fellow Cornellian Claire Cardie's startup, Jodange)
- At least 550 papers have referred to the topic since 2002 [Google scholar]

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

- Challenges: a few examples showing the complexity of the language phenomena involved
- Selected algorithmic ideas: domain adaptation, modeling label trajectories, and collective classification
- Connections to other areas: politics, economics, sociology/social psychology

Very restricted, "easy" case: polarity classification

Consider just classifying an avowedly subjective text unit as either positive or negative ("thumbs up or "thumbs down").

One application: review summarization.

Very restricted, "easy" case: polarity classification

Consider just classifying an avowedly subjective text unit as either positive or negative ("thumbs up or "thumbs down").

One application: review 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.*

Very restricted, "easy" case: polarity classification

Consider just classifying an avowedly subjective text unit as either positive or negative ("thumbs up or "thumbs down").
One application: review 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 more than a dozen others, at least]

Can't we just look for words like "great" or "terrible"?
Yes, but ...
... in a small-scale human study
[Pang, Lee, and 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%

Can't we just look for words like "great" or "terrible"?
Yes, but ...
... in a small-scale human study
[Pang, Lee, and 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%

Can't we just look for words like "great" or "terrible"?
Yes, but ...

- This laptop is a great deal.
- A great deal of media attention surrounded the release of the new laptop.
- This laptop is a great deal ... and I've got a nice bridge you might be interested in.

Can't we just look for words like "great" or "terrible"?
Yes, but ...

- This laptop is a great deal.
- A great deal of media attention surrounded the release of the new laptop.
- 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.

Can't we just look for words like "great" or "terrible"?
Yes, but ...

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

Can't we just look for words like "great" or "terrible"?
Yes, but ...

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

Indeed, "just" polarity classification has proven harder than topic classification [Pang, Lee, and Vaithyanathan '02].

We'll now look at a few particularly interesting algorithmic ideas that have been applied to sentiment analysis.
Theme: relationships (btwn. domains, labels, and/or items)

Supervised learning and the domain-dependence problem

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].

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

This dependence holds even though the set of categories (e.g., 1-5 stars) is often constant across domains.

Domain adaptation via structural correspondence learning

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

Book reviews (source domain: labeled)

(–) terrible and predictable

(–) absolutely terrible

(+) great

(–) so predictable it's terrible

Kitchen appliance reviews (target domain: unlabeled)

terrible leaking thing

truly terrible

OK

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)

(–) terrible and predictable

(–) absolutely terrible

(+) great

(–) so predictable it's terrible

Kitchen appliance reviews (target domain: unlabeled)

terrible leaking thing

truly terrible

OK

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)

(–) terrible and predictable

(–) absolutely terrible

(+) great

(–) so predictable it's terrible

Kitchen appliance reviews (target domain: unlabeled)

terrible leaking thing

truly terrible

OK

terrible, always leaking

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

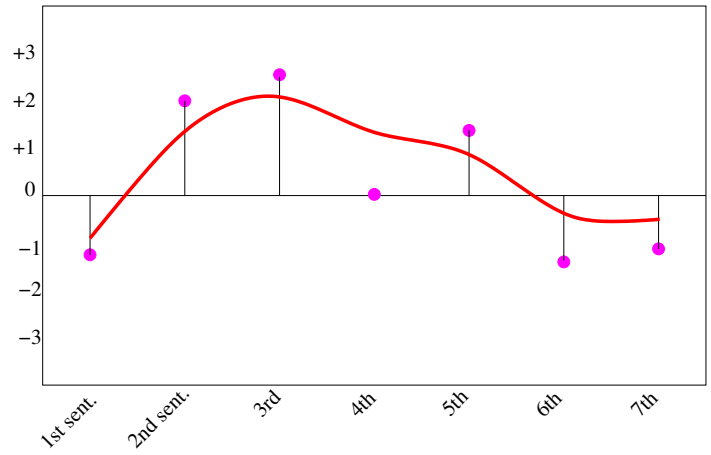
Relationships between sentiment labels

- "1 star" is more like "2 stars" than like "5 stars"
- "high" force of opinion is more like "medium" than "none"
- the "positive"-to-"negative" continuum may not contain the class "objective" ($\neq$ "mixed")

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] combine a sequence of local sentiment judgments into a (smoothed) 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.

This implements the intuition that a document can be represented by the "flow" of its discourse.

Sentiment analysis using discussion structure

Thomas, Pang, and Lee ['06]: classify speeches in US Congressional floor debates as supporting or opposing the proposed legislation.

Sentiment analysis using discussion structure

Thomas, Pang, and Lee ['06]: classify speeches in US Congressional floor debates as supporting or opposing the proposed legislation.

Two sources of information (details suppressed):

- An individual-document classifier that scores each speech in isolation
- An *agreement (degree) classifier* for *pairs* of speeches

Sentiment analysis using discussion structure

Thomas, Pang, and Lee ['06]: classify speeches in US Congressional floor debates as supporting or opposing the proposed legislation.

Two sources of information (details suppressed):

- An individual-document classifier that scores each speech in isolation
- An *agreement (degree) classifier* for *pairs* of speeches

Agreement info allows collective classification — "agreeing speeches should get the same label".

For certain formulations, one can *efficiently* produce an *optimal* solution via finding *minimum cuts* in the appropriate graph [see also Bansal, Cardie, and Lee '08].

Broader implications: politics

The on-line availability of politically-oriented documents, both official (e.g., parliamentary debates) and non-official (e.g., blogs), means:

The "[alteration of] the citizen-government relationship" [Shulman and Schlosberg 2002]
"The transformation of American politics" [*The New York Times*, 2006]
"The End of News?" [*The New York Review of Books*, 2005]

More opportunities for sentiment analysis!

Recall: people are searching for political news and perspectives.

Broader implications: politics

The on-line availability of politically-oriented documents, both official (e.g., parliamentary debates) and non-official (e.g., blogs), means:

The "[alteration of] the citizen-government relationship" [Shulman and Schlosberg 2002]
"The transformation of American politics" [*The New York Times*, 2006]
"The End of News?" [*The New York Review of Books*, 2005]

More opportunities for sentiment analysis!

Recall: people are searching for political news and perspectives.

One ought to recognize that the present political chaos is connected with the decay of language, and that one can probably bring about some improvement by starting at the verbal end.

Broader implications: politics

The on-line availability of politically-oriented documents, both official (e.g., parliamentary debates) and non-official (e.g., blogs), means:

The "[alteration of] the citizen-government relationship" [Shulman and Schlosberg 2002]
"The transformation of American politics" [*The New York Times*, 2006]
"The End of News?" [*The New York Review of Books*, 2005]

More opportunities for sentiment analysis!

Recall: people are searching for political news and perspectives.

> *One ought to recognize that the present political chaos is connected with the decay of language, and that one can probably bring about some improvement by starting at the verbal end.*
>
> — George Orwell, "Politics and the English language", 1946

Broader implications: economics

Consumers *report* being willing to pay from 20% to 99% more for a 5-star-rated item than a 4-star-rated item. [comScore]

But, does the polarity and/or volume of reviews have measurable, significant influence on actual consumer purchasing?

- ▷ Implications for bang-for-the-buck, manipulation, etc.

Broader implications: economics

Consumers *report* being willing to pay from 20% to 99% more for a 5-star-rated item than a 4-star-rated item. [comScore]

But, does the polarity and/or volume of reviews have measurable, significant influence on actual consumer purchasing?

▷ Implications for bang-for-the-buck, manipulation, etc.

Sample quote (much debate in the literature):

...on average, 3.46 percent of [eBay] sales is attributable to the seller's positive reputation stock. ... the average cost to sellers stemming from neutral or negative reputation scores is $2.28, or 0.93 percent of the final sales price.

Broader implications: economics

Consumers *report* being willing to pay from 20% to 99% more for a 5-star-rated item than a 4-star-rated item. [comScore]

But, does the polarity and/or volume of reviews have measurable, significant influence on actual consumer purchasing?

▷ Implications for bang-for-the-buck, manipulation, etc.

Sample quote (much debate in the literature):

...on average, 3.46 percent of [eBay] sales is attributable to the seller's positive reputation stock. ... the average cost to sellers stemming from neutral or negative reputation scores is $2.28, or 0.93 percent of the final sales price. **If these 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]*

Broader implications: sociology

What opinions are influential?

$\longrightarrow$ proxy question: which Amazon reviews are rated helpful?

[Danescu-Niculescu-Mizil, Kossinets, Kleinberg, and Lee '09]

Broader implications: sociology

What opinions are influential?

$\longrightarrow$ proxy question: which Amazon reviews are rated helpful?

[Danescu-Niculescu-Mizil, Kossinets, Kleinberg, and Lee '09]

Prior work has focused on features of the *text* of the reviews, and has not been in the context of sociological inquiry. [Kim et al. '06, Zhang and Varadarajan '06, Ghose and Ipeirotis '07, Jindal and B. Liu '07, J. Liu et al '07].

Our focus: how about *non-textual* features (social aspects, biases)?

Our corpus: millions of Amazon book reviews.

Some social factors boosting helpfulness scores

- using “real name”

Some social factors boosting helpfulness scores

- using “real name”
- being from New Jersey (for science books)

Some social factors boosting helpfulness scores

- using “real name”
- being from New Jersey (for science books)
- not being from Guam

Some social factors boosting helpfulness scores

- using "real name"
- being from New Jersey (for science books)
- not being from Guam

What about the review's star rating in relationship to others?

Theories from sociology/social psychology:

- conform (to the average rating) [Bond and Smith '96]
- "brilliant but cruel" [Amabile '83]

Some social factors boosting helpfulness scores

- using "real name"
- being from New Jersey (for science books)
- not being from Guam

What about the review's star rating in relationship to others?

Theories from sociology/social psychology:

- conform (to the average rating) [Bond and Smith '96]
- "brilliant but cruel" [Amabile '83]

As *variance* among reviews increases, be *slightly above* the mean

Some social factors boosting helpfulness scores

- using "real name"
- being from New Jersey (for science books)
- not being from Guam

What about the review's star rating in relationship to others?

Theories from sociology/social psychology:

- conform (to the average rating) [Bond and Smith '96]
- "brilliant but cruel" [Amabile '83]

As *variance* among reviews increases, be *slightly above* the mean ... except in Japan, where it's best to be *slightly below*.

Are the social effects just textual correlates?

We would like to control for the actual quality of a review's text. (Maybe people from NJ write inherently better reviews about science books?)

- manual annotation? Tedious, subjective.
- automatic classification? Need extremely high accuracy guarantees.

Are the social effects just textual correlates?

We would like to control for the actual quality of a review's text. (Maybe people from NJ write inherently better reviews about science books?)

- manual annotation? Tedious, subjective.
- automatic classification? Need extremely high accuracy guarantees.

It turns out that 1% of Amazon reviews are *plagiarized*! (see also David and Pinch ['06]).

Our social-effects findings regarding position relative to the mean hold on plagiarized pairs, which *by definition* have the same textual quality.

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/home/llee/opinion-mining-sentiment-analysis-survey.html
135 pp, 330+ references, pdf available online

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

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 (and students in your classes) 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 seen that sentiment analysis...

...has many important applications
...encompasses many interesting research questions
...extends to many areas

You (and students in your classes) 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!