Introduction to Sentiment Analysis

School of Information Studies
Syracuse University

Why Sentiment Analysis?

Movie: Is this review positive or negative?

Products: What do people think about the new iPhone?

Customer relations: What do people say about your company?

Public sentiment: How is consumer confidence?
Is despair increasing?

Politics: What do people think about this candidate
or issue?

Prediction: Predict election outcomes or market trends from sentiment.

Types of Tasks

Positive or negative movie review?

- *Unbelievably disappointing*

- *Full of zany characters and richly applied satire, and some great plot twists*

- *This is the greatest screwball comedy ever filmed.*

- *It was pathetic. The worst part about it was the boxing scenes.*

Types of Tasks: Google Product Search

HP Officejet 6500A Plus e-All-in-One Color Ink-jet - Fax / copier / printer / scanner
$89 online, **$100** nearby ★★★★☆ 377 reviews
September 2010 - Printer - HP - Inkjet - Office - Copier - Color - Scanner - Fax - 250 sh

Reviews

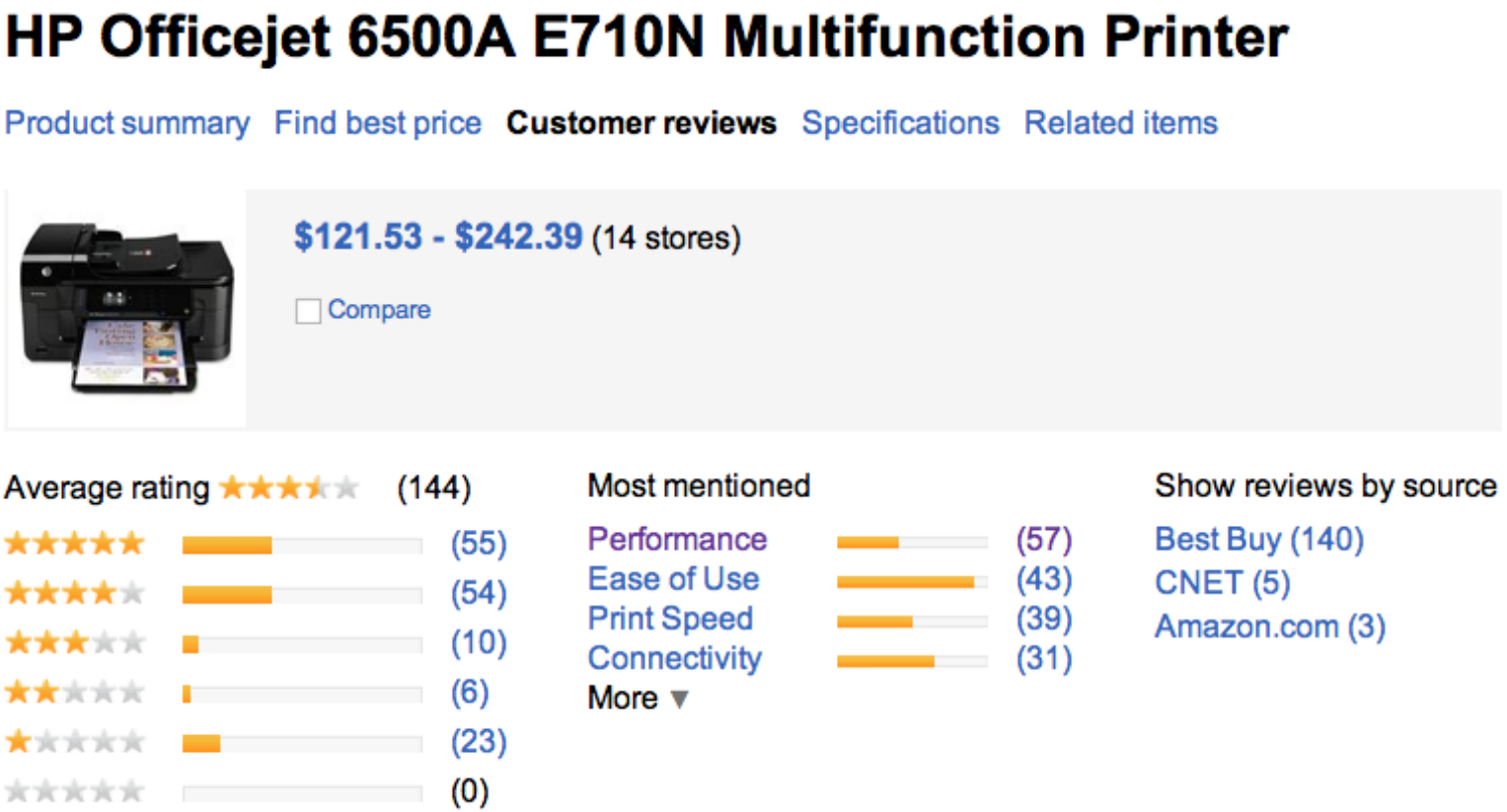
Summary - Based on 377 reviews

1 star	2	3	4 stars	5 stars

What people are saying

ease of use	"This was very easy to setup to four computers."
value	"Appreciate good quality at a fair price."
setup	"Overall pretty easy setup."
customer service	"I DO like honest tech support people."
size	"Pretty Paper weight."
mode	"Photos were fair on the high quality mode."
colors	"Full color prints came out with great quality."

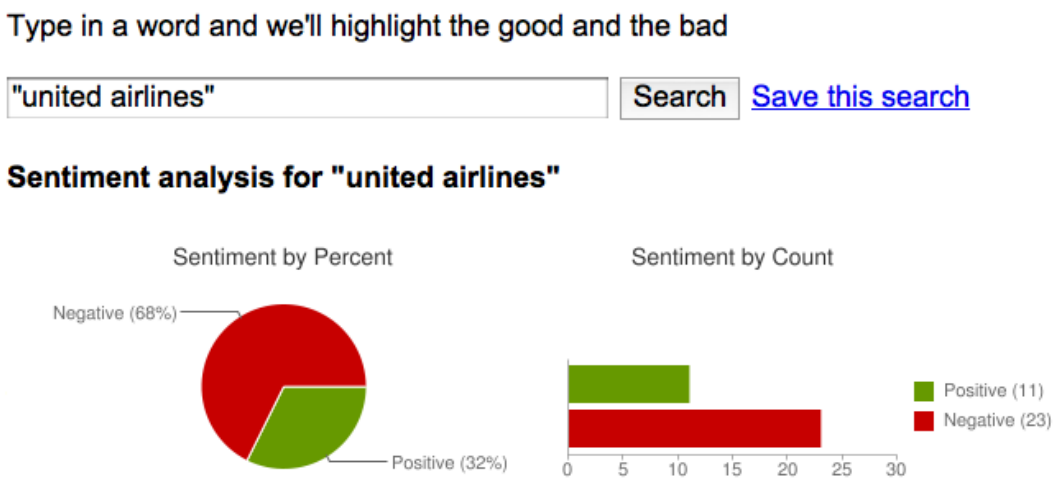
Types of Tasks: Bing Shopping



Target Sentiment on Twitter

Twitter Sentiment App

Alec Go, Richa Bhayani, & Le Huang. (2009). Twitter Sentiment Classification Using Distant Supervision



jljacobson: OMG... Could **@United airlines** have worse customer service? W8g now 15 minutes on hold 4 questions about a flight 2DAY that need a human.
Posted 2 hours ago

12345clumsy6789: I hate **United Airlines** Ceiling!!! Fukn impossible to get my conduit in this damn mess! ?
Posted 2 hours ago

EMLandPRGbelgiu: EML/PRG fly with Q8 **united airlines** and 24seven to an exotic destination. http://t.co/Z9QloAjF
Posted 2 hours ago

CountAdam: FANTASTIC customer service from **United Airlines** at XNA today. Is tweet more, but cell phones off now!
Posted 4 hours ago

Definitions for Sentiment Analysis

School of Information Studies
Syracuse University

Terminology

For the more subtle aspects of the semantics of text:

Affective aspects of text is that which is "influenced by or resulting from emotions"

- One aspect of non-factual aspects of text

Subjective aspects of text are "the linguistic expression of somebody's opinions, sentiments, emotions, evaluations, beliefs, speculations (private states)"

- A private state is not open to objective observation or verification
- Subjectivity analysis would classify parts of text as to whether it was subjective or objective

Scherer Typology of Affective States

Emotion: brief organically synchronized… evaluation of a major event, emotion is subject to change

- *angry, sad, joyful, fearful, ashamed, proud, elated*

Mood: diffuse, non-caused, low-intensity, long-duration change in subjective feeling

- *cheerful, gloomy, irritable, listless, depressed, buoyant*

Interpersonal stances: affective stance toward another person in a specific interaction

- *friendly, flirtatious, distant, cold, warm, supportive, contemptuous*

Attitudes: enduring, affectively colored beliefs, dispositions toward objects or persons

- *liking, loving, hating, valuing, desiring*

Personality traits: stable personality dispositions and typical behavior tendencies

- *nervous, anxious, reckless, morose, hostile, jealous*

Sentiment Analysis Focuses on Attitudes

Emotion: brief organically synchronized… evaluation of a major event, emotion is subject to change

- *angry, sad, joyful, fearful, ashamed, proud, elated*

Mood: diffuse, non-caused, low-intensity, long-duration change in subjective feeling

- *cheerful, gloomy, irritable, listless, depressed, buoyant*

Interpersonal stances: affective stance toward another person in a specific interaction

- *friendly, flirtatious, distant, cold, warm, supportive, contemptuous*

Attitudes: enduring, affectively colored beliefs, dispositions toward objects or persons

- *liking, loving, hating, valuing, desiring*

Personality traits: stable personality dispositions and typical behavior tendencies

- *nervous, anxious, reckless, morose, hostile, jealous*

Sentiment Analysis

Sentiment analysis is the detection of attitudes

- "Enduring, affectively colored beliefs, dispositions toward objects or persons"
- **Holder (source)** of attitude
- **Target (aspect)** of attitude
- **Type of attitude**
  - From a set of types
    - Like, love, hate, value, desire, etc.
  - Or (more commonly) simple weighted polarity
    - Positive, negative, neutral, together with strength
- **Text** containing the attitude
  - Sentence or entire document

Sentiment Analysis Task Levels

Simplest task is polarity

- Is the attitude of this text positive or negative?
  - Negative/positive attitude of reporter/blogger
  - Favorable/unfavorable review of a product
  - Right/left political leaning of speaker

More complex

- Rank the attitude of this text from 1 to 5
  - Sometimes called strength or intensity

Advanced

- Detect the target, source, or complex attitude type
- May also be referred to as opinion extraction, opinion mining, or sentiment mining

What's the Problem?

Consider classifying a subjective text unit as either positive or negative.

- Example: the most thoroughly joyless and inept film of the year, and one of the worst of the decade [Mick LaSalle, describing *Gigli*]

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.
- Words have different meanings in different contexts

What's the Problem?

Subtlety, sarcasm, or metaphor

- Perfume review in *Perfumes: The Guide*
  - *"If you are reading this because it is your darling fragrance, please wear it at home exclusively, and tape the windows shut."*
- Dorothy Parker on Katherine Hepburn
  - *"She runs the gamut of emotions from A to B."*

Ordering Effects

Thwarted expectations

- A lot of good words set up an expectation that is then negated.
  - *"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."*
- Or general expectations are negated.
  - *Well, as usual, Keanu Reeves is nothing special, but surprisingly, the very talented Laurence Fishburne is not so good either. I was surprised.*

Domain Adaptation

Certain sentiment-related indicators seem domain-dependent.

- Read the book—good for book reviews, bad for movie reviews
- Unpredictable—good for movie plots, bad for a car's steering [Turney '02]

In general, sentiment classifiers (especially those created via supervised learning) **have been shown to often be domain dependent.**

- Classifiers trained on one domain—for example, movies—do not do well in another domain—for example, other products or political sentiment

Sentiment Analysis: Approach and Negation

School of Information Studies
Syracuse University

Sentiment Polarity

Classic sentiment polarity task from Pang and Lee

- Is an IMDB movie review positive or negative?
- Data: Polarity Data 2.0: (people indicate polarity of own review)
  - http://www.cs.cornell.edu/people/pabo/movie-review-data

Treat as a document classification task

- Positive, negative, and (possibly) neutral

Similar but different from topic-based text classification

- In topic-based text classification, topic words are important
- In sentiment classification, sentiment words are more important (e.g., *great, excellent, horrible, bad, worst,* etc.)

IMDB Data in the Movie Corpus

Examples of how people discuss movies in reviews.

✓

when _star wars_ came out some twenty years ago , the image of traveling throughout the stars has become a commonplace image. [...]

when han solo goes light speed , the stars change to bright lines , going towards the viewer in lines that converge at an invisible point.

cool.

october sky offers a much simpler image–that of a single white dot , traveling horizontally across the night sky. [...]

✗

"snake eyes " is the most aggravating kind of movie : the kind that shows so much potential then becomes unbelievably disappointing .

it's not just because this is a brian depalma film , and since he's a great director and one who's films are always greeted with at least some fanfare .

and it's not even because this was a film starring nicolas cage and since he gives a bravura performance , this film is hardly worth his talents .

Treat as a Classification Problem

Tokenization

- May be some differences, especially for text from social media

Feature extraction

- The most important part!

Classification using different classifiers

- Naïve Bayes
- MaxEnt
- SVM
  - It turns out that MaxEnt and SVM are better than Naïve Bayes at some sentiment domains.

Sentiment Tokenization Issues

For text from Web, deal with HTML and XML markup

Or Twitter mark-up (names, hashtags)

Capitalization (preserve for words in all caps)

Phone numbers, dates

Emoticons

Useful code for twitter and other social media text:

- Christopher Potts sentiment tokenizer
- Brendan O'Connor Twitter tokenizer
  - Now in NLTK

Potts emoticons

```
[<>]?                # optional hat/brow
[:;=8]               # eyes
[\-o\*\']?           # optional nose
[\)\]\(\[dDpP/\:\}\{@\|\\]    # mouth
|                   ## reverse
orientation
[\)\]\(\[dDpP/\:\}\{@\|\\]    # mouth
[\-o\*\']?           # optional nose
[:;=8]               # eyes
[<>]?                # optional hat/brow
```

Extracting Features for Sentiment Classification

Which words to use?

- Only adjectives
- Or, all words
  - All words turn out to work better, at least on these data

Syntax

- Counts of POS tags to characterize text
- Constituent or dependency parses
  - Particularly at phrase level to find dependencies of opinion words
  - Also for finding the scope of negation

Handling Negation Is Important!

How to handle negation

- *I didn't like this movie* vs. *I really like this movie*

Typical approach

- Given a "prototype" negation word like *not, no,* and *never*
- Add a "negated context" to the features
  - Pang and Lee simple approximation in this approach assuming that the scope of negated word goes until next punctuation
  - Add **not_** to every word between negation and following punctuation in the features:

didn't like this movie , but I

didn't **not_**like **not_**this **not_**movie but I

Negation

Negation has both cue words and scope

- **Need to identify the negation cue words**
  - In addition to prototype negators (words that always indicate negation), there are approximate negators that sometimes indicate full negation (e.g., *hardly, scarcely,* etc.)
  - In the research literature, lists of negation cue words seldom agree
- **Scope can primarily be identified from syntax**
  - Although this is sometimes difficult in social media text due to lack of standard punctuation

Degrees of Negation

Current research focus on how positive or negative sentiment can be:

- Sometimes scaled from –5 to +5

The negation cues and scope can also have degrees, sometimes called **valence**

- Some words, typically adverbs or adjectives, can be valence shifters
  - For example, intensification and diminution
    - *Very good, deeply suspicious*
    - *Should have been good*
  - May also shift to the opposite
    - *He is a great actor; however, this performance…*
      - The word "however" changes the valence of "great" to be negative

Sentiment Analysis: Lexicons

School of Information Studies
Syracuse University

Sentiment Lexicons

One of the early approaches to sentiment analysis was to just count the words in each document that had either a positive or negative polarity from a (hand-built) sentiment lexicon.

- This approach is usually not very accurate on individual documents, but it's easy because it doesn't need training data.
- It may be useful over aggregate collections or to show trends over time.

Now we use **either presence or frequencies of sentiment words as features of the classifier.**

- This is an important part of the sentiment analysis approach.
- Features are usually included from one or more sentiment lexicons.
  - Sometimes both standard lexicons and specially built domain lexicons are used.

MPQA Subjectivity Cues Lexicon

This lexicon contains a list of words that have been judged to be weakly or strongly positive, negative, or neutral in subjectivity

Home page: http://www.cs.pitt.edu/mpqa/subj_lexicon.html

6,885 words from 8,221 lemmas

- 2,718 positive, 4,912 negative
- Examples excerpted from the file:

```
type=weaksubj word1=abandoned pos1=adj stemmed1=n
priorpolarity=negative
type=strongsubj word1=absolve pos1=verb stemmed1=y
priorpolarity=positive
type=strongsubj word1=absolute pos1=adj stemmed1=n
priorpolarity=neutral
```

T. Wilson, J. Wiebe, & P. Hoffmann (2005). Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. Proc. of HLT-EMNLP-2005.

LIWC (Linguistic Inquiry and Word Count)

Linguistic Inquiry and Word Count

- Text analysis software based on dictionaries of word dimensions
- Dimensions can be syntactic
  - Pronouns, past-tense verbs
- Dimensions can be semantic
  - Social words, affect, cognitive mechanisms
- Other categories
  - See http://www.liwc.net/comparedicts.php

James Pennebaker, University of Texas at Austin

- $30–90 fee for software (make sure to get dictionaries)

Often used for positive and negative emotion words in opinion mining

ANEW (Affective Norms for English Words)

Provides a set of emotional ratings for a large number of words in the English language

Participants gave graded reactions, 1–9, on three dimensions

- Good/bad, psychological valence
- Active/passive, arousal valence
- Strong/weak, dominance valence

From the NIMH Center for the Study of Emotion and Attention at the University of Florida

- http://csea.phhp.ufl.edu/Media.html
- See also the paper by Dodds and Danforth on Happiness of Large-Scale Written Expressions
- Free for research use

The General Inquirer

- Home page: http://www.wjh.harvard.edu/~inquirer
- List of categories: http://www.wjh.harvard.edu/~inquirer/homecat.htm
- Spreadsheet: http://www.wjh.harvard.edu/~inquirer/inquirerbasic.xls

Categories

- Positiv (1,915 words) and Negativ (2,291 words)
- Strong vs. weak, active vs. passive, overstated vs. understated
- Pleasure, pain, virtue, vice, motivation, cognitive orientation, etc.

Free for research use

P. J. Stone, D. C. Dunphy, M. S. Smith, & D. M. Ogilvie. (1966). *The General Inquirer: A Computer Approach to Content Analysis.* MIT Press.

Bing Liu Opinion Lexicon

Bing Liu's Page on Opinion Mining

http://www.cs.uic.edu/~liub/FBS/opinion-lexicon-English.rar

6,786 words

- 2,006 positive
- 4,783 negative

M. Hu & B. Liu. (2004). Mining and Summarizing Customer Reviews. ACM SIGKDD-2004.

SentiWordNet

- Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. (2010). *SENTIWORDNET 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining.* LREC-2010
- Home page: http://sentiwordnet.isti.cnr.it/
- All WordNet synsets automatically annotated for degrees of positivity, negativity, and neutrality/objectiveness
- [estimable(J,3)] "may be computed or estimated"
  - Pos 0 Neg 0 Obj 1
- [estimable(J,1)] "deserving of respect or high regard"
  - Pos .75 Neg 0 Obj .25

Which Sentiment Lexicon to Use?

An area of active research in the sentiment-analysis community

It is now recognized that the amount of overlap between the lexicons is small!

- But, in general, where there is overlap, the sentiment polarity of the words is in agreement, 2% or less disagreement
  - Except for SentiWordNet, which disagrees up to 25%
  - Chris Potts, Sentiment Symposium Tutorial

How to represent features from sentiment words still under research

- Frequency of all positive and all negative words
- Presence of positive or negative words (particularly for twitter)
- Sum of the positive or negative intensity scores

Building a Sentiment Lexicon

For some domains, it has been shown that the best lexicon is one built for that domain

Automatic lexicon building from unlabeled data

- Bootstrapping
  - Identify a number of seed words of positive and negative polarity
  - Search for text involving those words that also has connecting words, such as “and”
  - Other words that occur with the connecting word are added to the lexicon with the appropriate polarity
- Trained from annotated text by associating words with the sentiment labels with which they occur
  - Using Mutual Information scores or other measures

Building a Sentiment Lexicon

Automatic lexicon building from labeled data

- In some cases, the domain has a lot of text that has been labeled with sentiment
- Twitter
  - Use tweets labeled with sentiment hashtags: #good, #happy, #bad, #sad
  - Use tweets labeled with happy or sad emoticons
- Collect words from the positive and negative labeled texts and keep the frequent ones as part of a lexicon

Sentiment Analysis: Advanced Topics

School of Information Studies
Syracuse University

Latest Sentiment Detection on Sentences

The most recent advances in sentiment detection combine the sentiment of phrases and partial phrases to get the overall sentiment for a sentence

Needs sentiment for the sub-phrases

- May use sentiment phrase labeling: see the kaggle dataset of phrase labeling of movie reviews

Needs a method to combine the sentiment

- Can use classification
- Or deep-learning neural nets

Movie Review Sentences

Results on binary (pos/neg)

- Straight classification on sentences: under 80%
- Richard Socher—combining phrases by deep learning: 85%

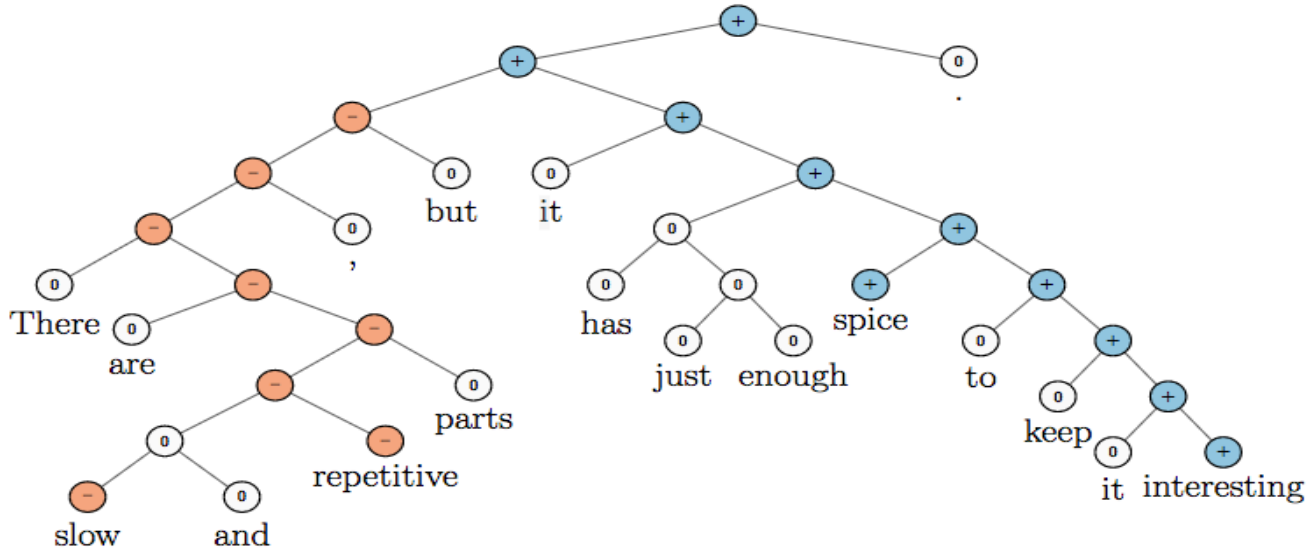


Figure 7: Example of correct prediction for contrastive conjunction *X but Y*.

How to Deal With Seven Stars?

The second level of sentiment analysis deals is a similar classification task but needs to find levels of strength

The typical approach is to use machine-learning algorithms that can give numbers of probability or certainty (regression tasks).

- For example, can be used to classify with 1 to 7
- Uses the same or similar features as the polarity task

B. Pang & L. Lee. (2005). Seeing Stars: Exploiting Class Relationships for Sentiment Categorization With Respect to Rating Scales. Proc. of ACL, 115–124.

Opinion Mining

School of Information Studies
Syracuse University

Opinion Mining

The **third level of sentiment analysis** is sometimes called opinion mining because you are finding sentiment toward aspects or attributes.

Businesses spend a huge amount of money to find consumer sentiments and opinions.

- Consultants, surveys and focus groups, etc.
- Text in the form of transcripts of interviews or survey responses

Opinions are also available on the Web.

- Product reviews
- Blogs, discussion groups

Sentence Level Detection

Sentence level or sub-sentence level detection of subjectivity as a binary classifier

- Wiebe, many projects
- Pang and Lee: for movie reviews, first determine which sentences express opinions and then label for opinion polarity

Clause level opinion strength

- Wilson: "How mad are you?"

Detection of sentences with subjectivity or sentiment is important to then find aspects or attributes and not just overall sentiment

- *The food was great but the service was awful.*

Finding Aspect/Attribute/Target of Sentiment

Frequent phrases + rules

- Find all highly frequent phrases across reviews ("fish tacos")
- Filter by rules like "occurs right after sentiment word"
  - "…great fish tacos" means fish tacos a likely aspect

Casino	casino, buffet, pool, resort, beds
Children's Barber	haircut, job, experience, kids
Greek Restaurant	food, wine, service, appetizer, lamb
Department Store	selection, department, sales, shop, clothing

M. Hu & B. Liu. (2004). Mining and Summarizing Customer Reviews. Proc. of KDD.
S. Blair-Goldensohn, K. Hannan, R. McDonald, T. Neylon, G. Reis, & J. Reynar. (2008). Building a Sentiment Summarizer for Local Service Reviews. WWW Workshop.

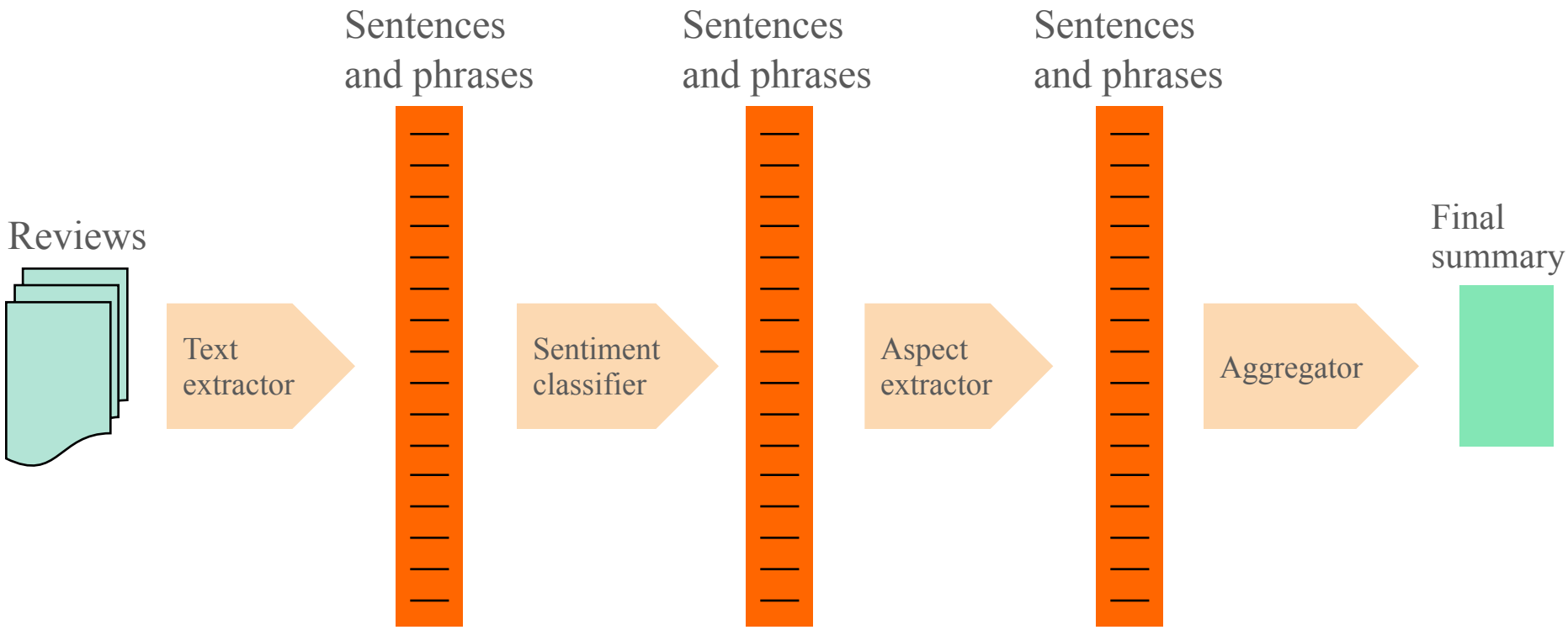
Finding Aspect/Attribute/Target of Sentiment

But the aspect name may not be in the sentence

Another method to find aspects

- For restaurants/hotels, aspects are well understood
- Use those aspects in supervised classification
  - Hand-label a small corpus of restaurant review sentences with aspect
  - food, décor, service, value, **none**
  - Train a classifier to assign an aspect to a sentence
  - "Given this sentence, is the aspect food, décor, service, value, or **none**"

Putting It All Together: Finding Sentiment/Opinion for Aspects



S. Blair-Goldensohn, K. Hannan, R. McDonald, T. Neylon, G. Reis, & J. Reynar. (2008). Building a Sentiment Summarizer for Local Service Reviews. WWW Workshop.

Joint Topic/Sentiment Analysis

An alternative approach to first finding the aspect or attribute and then the opinion or sentiment is to find them both in the same classification

- Comparative studies of related products
- Topics that have various features and attributes
  - Consumers
  - Political areas

Example Results for Opinion on Aspects

Rooms (3/5 stars, 41 comments)

- (+) The room was clean and everything worked fine, even the water pressure…
- (+) We went because of the free room and was pleasantly pleased…
- (–) …the worst hotel I had ever stayed at…

Service (3/5 stars, 31 comments)

- (+) Upon checking out, another couple was checking early due to a problem…
- (+) Every single hotel staff member treated us great and answered every…
- (–) The food is cold and the service gives new meaning to **slow**…

Dining (3/5 stars, 18 comments)

- (+) Our favorite place to stay in Biloxi. The food is great also the service…
- (+) Offer of free buffet for joining the play…

Feature-Based Summary

Companies like Amazon use feature-based summary techniques to summarize and present product reviews.

GREAT Camera., Jun 3, 2004
Reviewer: **jprice174** from Atlanta, Ga.

> I did a lot of research last year before I bought this camera... It kinda hurt to leave behind my beloved nikon 35mm SLR, but I was going to Italy, and I needed something smaller, and digital.
>
> The pictures coming out of this camera are amazing. The 'auto' feature takes great pictures most of the time. And with digital, you're not wasting film if the picture doesn't come out. …

….

Feature-Based Summary:
Feature1: picture
Positive: 12

- The pictures coming out of this camera are amazing.
- Overall this is a good camera with a really good picture clarity.

…
Negative: 2

- The pictures come out hazy if your hands shake even for a moment during the entire process of taking a picture.
- Focusing on a display rack about 20 feet away in a brightly lit room during day time, pictures produced by this camera were blurry and in a shade of orange.

Feature2: **battery life**
…

Other Tasks Related to Sentiment Analysis

School of Information Studies
Syracuse University

Other Tasks: Subjectivity Detection

For many applications, first decide if the document contains subjective information or which parts are subjective.

- Focus of TREC 2006 Blog track.
- Label sentences or documents as to whether they are objective or subjective.

Computational Work on Other Affective States

Emotion

- Detecting annoyed callers to dialogue system
- Detecting confused/frustrated vs. confident students

Mood

- Finding traumatized or depressed writers
  - One study used transcripts from a suicide hotline to try to detect possible suicides for intervention

Interpersonal stances

- Detection of flirtation or friendliness in conversations

Personality traits

- Detection of extroverts

Certainty

- How certain the holder is of the opinion or sentiment

Detection of Friendliness

After speed dating conversations, students rated their conversational partners as to friendliness

Friendly speakers use collaborative conversational style

- Laughter
- Less use of negative emotional words
- More sympathy
  - *That's too bad. I'm sorry to hear that.*
- More agreement
  - *I think so too.*
- Fewer hedges
  - *kind of, sort of, a little…*

Ranganath, Jurafsky, McFarland

Viewpoints and Perspectives

In some types of documents, the authors are not necessarily discussing opinions on particular topics but are revealing general attitudes or sometimes a set of bundled attitudes and beliefs.

- Classifying political blogs as liberal, conservative, libertarian, etc.
- Identifying Israeli vs. Palestinian viewpoints

One type of this is multi-perspective question answering

- On next slide…

MPQA

Multi-perspective question answering (MPQA)

- *What does Bush think about Hillary Clinton?*
- *How does the United States regard the latest terrorist attacks in Baghdad?*

Sentence, or part of a sentence, that answers the question:

- *"How does X feel about Y?"*
- *"It makes the system more flexible," argues a Japanese businessman.*

Looking for opinion linked to opinion holder

- Where the opinion holder is someone other than the author

Stoyanov, Cardie, Wiebe, & Litman. (2004). **Evaluating an Opinion Annotation Scheme Using a Multi-Perspective Question and Answer Corpus.** 2004 AAAI Spring Symposium on Exploring Attitude and Affect in Text.

Stance and Argumentation

Some forms of online discourse takes the form of trying to argue a viewpoint or opinion, or taking a stance in a particular debate.

- **Ideological debates**
  - Somasundaram and Wiebe: look at argumentation
  - Abbot, Walker, et al.: classifying stance in online debates
    - The "Cats rule, dogs drool!"debate is much easier to classify than debates on abortion, religion, politics
    - We think people express opinions on less serious topics using more straightforward language
    - But use subtleties of expression or "cue phrases" on sensitive topics
- **Approaches use discourse markers of rhetorical structures** in addition to classification techniques

Sentiment Analysis: Two Examples

School of Information Studies
Syracuse University

Twitter Sentiment Detection

Twitter sentiment detection task is usually to label the entire tweet as positive, negative, or neutral in overall sentiment

Many efforts in this area, notably:

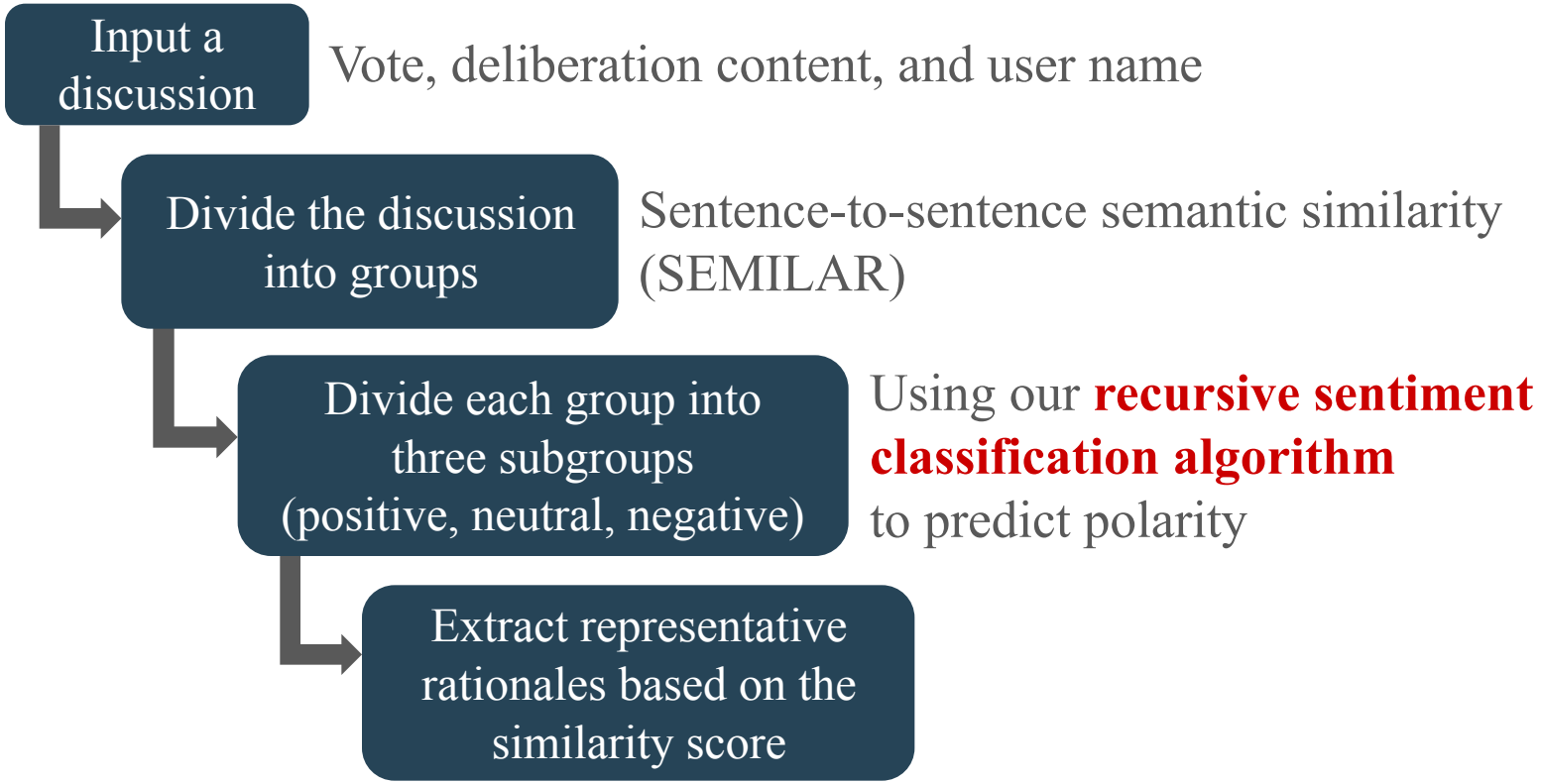
- Several years of shared tasks at the Semantic Evaluation Workshop (SemEval),2013–2017
- Details from a top performance by NRC-Canada-2014: Recent Improvements in Sentiment Analysis of Tweets, Xiaodan Zhu, Svetlana Kiritchenko, and Saif M. Mohammad
- Released description of features
  - Released resources created (tweet-specific sentiment lexicons)
- Importance of manually created sentiment lexicons and also tweet-specific automatically created lexicons

Examples of Twitter Sentiment Features

There were about 70,000 features in their classifier for each tweet.

Features	*Examples*
Sentiment lexicons	**#positive: 3, scorePositive: 2.2; maxPositive: 1.3; last: 0.6, scoreNegative: 0.8, scorePositive_neg: 0.4**
Word *N*-grams	**spectacular, like documentary**
Character *N*-grams	**spect, docu, visua**
Part of speech	**#N: 5, #V: 2, #A:1**
Negation	**#Neg: 1; ngram:perfect → ngram:perfect_neg, polarity:positive → polarity:positive_neg**
Word clusters	**probably, definitely, def**
All-caps	**YES, COOL**
Punctuation	**#!+: 1, #?+: 0, #!?+: 0**
Emoticons	**:D, >:(**
Elongated words	**soooo, yaayyy**

Detection of Representative Rationales in Wikipedia Deliberations



Slides from Professor Lu Xiao: (Mao, Xiao, & Mercer, 2014)

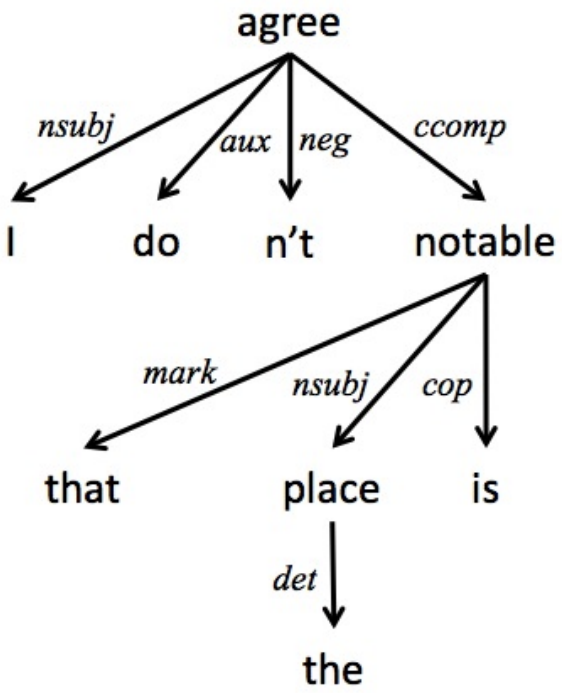
Recursive Algorithm to Predict Sentiment Polarity at Sentence Level

Based on the dependency structure tree

- Take a node as input and the polarity score for the node as output

The algorithm assigns a polarity score to each node in the dependency structure tree

Integrate five types of negations



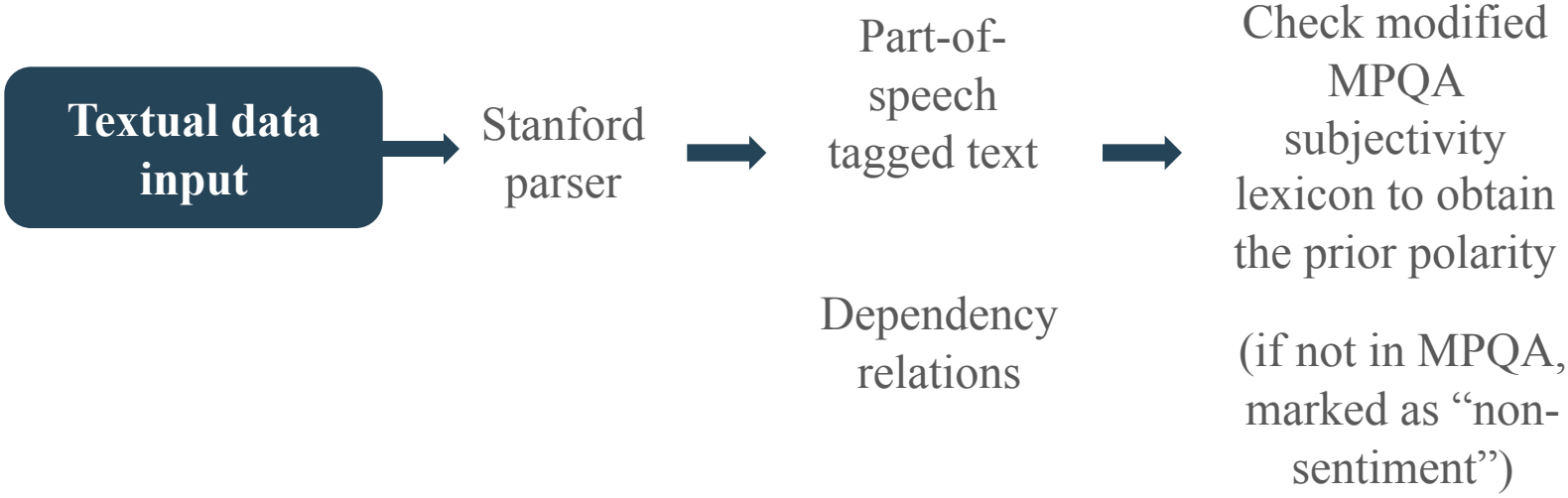
Five Negations

1. **Local negation:** a "not" usually modifies the sentiment word
   - "The place is *not notable*."
2. **Predicate negation:** using verbs with negative polarity
   - "I *disagree* that the place is notable."
3. **Subject negation:** a subject leads to the negation of its predicate
   - "*Neither* one of us *agrees* that the place is *notable*."
4. **Preposition negation:** the polarity of the object following the preposition "of" can be changed by the word modified by the preposition
   - "It is a *violation of notability*."
5. **Modifier negation:** some sentiment word's polarity can be negated by its modifier
   - "The place is of *indeterminable notability*."

Sentiment Analysis Module

Determine the sentiment polarity of a statement in our language context ("notable").

- MPQA subjectivity lexicon + additional words



Results for Sentence Polarity Prediction

Methods

- Stanford sentiment analysis tool vs. our algorithm
- Used 237 sentences from Wikipedia AfD discussions
  - Annotated with three classes: positive, neutral, and negative

	Stanford Sentiment Analysis	Recursive Algorithm Without Machine Learning	Recursive Algorithm With Machine Learning
Accuracy (%)	48.73	58.47	60.17

Accuracy of Stanford sentiment analysis tool classifying movie review

- Five-category: 45.7%; two-category: 85.4%

Slides from Professor Lu Xiao

Summary on Sentiment

Understanding semantics of less factual aspects of text

Sentiment analyzers available

- Vary widely as to what types of text and domains they work well on

Generally modeled as classification or regression task

- Predict a binary label for polarity
- Or, predict an ordinal label for the level of sentiment

Features

- Negation is important
- Using all (filtered) words works well for some tasks
- Finding sub-sets of words may help in other tasks
  - Hand-built polarity lexicons
  - Use seeds and semi-supervised learning to induce lexicons