SENTIMENT, OPINIONS, EMOTIONS

Heng Ji
jih@rpi.edu
April 2, 2014

Acknowledgement: Some slides from Jan Wiebe and Kavita Ganesan
OUTLINE

• Emotion Detection
• Subjectivity Overview
• Sentiment Analysis
• Opinion Mining
Emotion Examples

- A Happy Song?
  - [http://www.9ku.com/play/86601.htm](http://www.9ku.com/play/86601.htm)

- A Sad Song?
  - [http://www.9ku.com/play/197326.htm](http://www.9ku.com/play/197326.htm)

- Hard to draw the boundary…also depends on the audience’ mood
Emotion Clues

• Speech/Sound
• Text (lyrics)
• Face Expressions
• Comments @ Social Networks
Emotion Detection from Speech (Shriberg et al., 2001)

- Prosody = rhythm, melody, “tone” of speech
- Largely unused in current ASU systems
- Prior work: prosody aids many tasks:
  - Automatic punctuation
  - Topic segmentation
  - Word recognition
- Today’s talk: detection of user frustration in DARPA Communicator data (ROAR project suggested by Jim Bass)
Data Labeling

- **Emotion**: neutral, annoyed, frustrated, tired/disappointed, amused/surprised, no-speech/NA
- **Speaking style**: hyperarticulation, perceived pausing between words or syllables, raised voice
- **Repeats and corrections**: repeat/rephrase, repeat/rephrase with correction, correction only
- **Miscellaneous useful events**: self-talk, noise, non-native speaker, speaker switches, etc.
Prosodic Features

- **Duration and speaking rate features**
  - duration of phones, vowels, syllables
  - normalized by phone/vowel means in training data
  - normalized by speaker (all utterances, first 5 only)
  - speaking rate (vowels/time)

- **Pause features**
  - duration and count of utterance-internal pauses at various threshold durations
  - ratio of speech frames to total utt-internal frames
Features (cont.)

• **Spectral tilt features**
  - average of 1st cepstral coefficient
  - average slope of linear fit to magnitude spectrum
  - difference in log energies btw high and low bands
  - extracted from longest normalized vowel region

• **Other (nonprosodic) features**
  - position of utterance in dialog
  - whether utterance is a repeat or correction
  - to check correlations: hand-coded style features including hyperarticulation
Language Model Features

• Train 3-gram LM on data from each class
• LM used word classes (AIRLINE, CITY, etc.) from SRI Communicator recognizer
• Given a test utterance, chose class that has highest LM likelihood (assumes equal priors)
• In prosodic decision tree, use sign of the likelihood difference as input feature
• Finer-grained LM scores cause overtraining
## Results: Human and Machine

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy (%) (chance = 50%)</th>
<th>Kappa (Acc-C)/(1-C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Each Human with Other Human, overall</td>
<td>71.7</td>
<td>.38</td>
</tr>
<tr>
<td>Human with Human “Consensus” (biased)</td>
<td>84.2</td>
<td>.68</td>
</tr>
<tr>
<td>Prosodic Decision Tree with Consensus</td>
<td>75.6</td>
<td>.51</td>
</tr>
<tr>
<td>Tree with Consensus, no repeat/correction</td>
<td>72.9</td>
<td>.46</td>
</tr>
<tr>
<td>Tree with Consensus, repeat/correction only</td>
<td>68.7</td>
<td>.37</td>
</tr>
<tr>
<td>Language Model features only</td>
<td>63.8</td>
<td>.28</td>
</tr>
</tbody>
</table>
Hybrid Approach (Meghjani, 2011)

- Automatic emotion recognition using audio-visual information analysis.
- Create video summaries by automatically labeling the emotions in a video sequence.
Motivation

- Map Emotional States of the Patient to Nursing Interventions.
- Evaluate the role of Nursing Interventions for improvement in patient’s health.
Proposed Approach

- Visual Feature Extraction
- Audio Feature Extraction

- Visual based Emotion Classification
- Audio based Emotion Classification

- Decision Level Fusion
- Feature Level Fusion

- Data Fusion

- Recognized Emotional State
## Experimental Results

<table>
<thead>
<tr>
<th>Statistics Database</th>
<th>No. of Training Examples</th>
<th>No. of Subjects</th>
<th>No. of Emotional State</th>
<th>% Recognition Rate</th>
<th>Validation Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posed Visual Data Only (CKDB)</td>
<td>120</td>
<td>20</td>
<td>5+Neutral</td>
<td>75%</td>
<td>Leave one subject out cross validation</td>
</tr>
<tr>
<td>Posed Audio Visual Data (EDB)</td>
<td>270</td>
<td>9</td>
<td>6</td>
<td>82%</td>
<td>Decision Level</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>76%</td>
<td>Feature Level</td>
</tr>
</tbody>
</table>
VITALITY

• Emotion Detection
• Subjectivity Overview
• Sentiment Analysis
• Opinion Mining
“What people think?”

What others think has always been an important piece of information.

“What car should I buy?”

“What schools should I apply to?”

“What Professor to work for?”

“What should I vote for?”
“So whom shall I ask?”

Pre Web
• Friends and relatives
• Acquaintances
• Consumer Reports

Post Web
“I don’t know who..but apparently it’s a good phone. It has good battery life and…”
• Blogs (google blogs, livejournal)
• E-commerce sites (amazon, ebay)
• Review sites (CNET, PC Magazine)
• Discussion forums (forums.craigslist.org, forums.macrumors.com)
• Friends and Relatives (occasionally)
“Whoala! I have the reviews I need”

Now that I have “too much” information on one topic…I could easily form my opinion and make decisions…

Is this true?
…Not Quite

- Searching for reviews may be difficult
  - Can you search for opinions as conveniently as general Web search?
    eg: is it easy to search for “iPhone vs Google Phone”? 

- Overwhelming amounts of information on one topic
  - Difficult to analyze each and every review
  - Reviews are expressed in different ways
    “the google phone is a disappointment....”
    “don’t waste your money on the g-phone....”
    “google phone is great but I expected more in terms of...”
    “…bought google phone thinking that it would be useful but...”
“Let me look at reviews on one site only…”

Problems?

Biased views
- all reviewers on one site may have the same opinion

Fake reviews/Spam (sites like YellowPages, CitySearch are prone to this)
- people post good reviews about their own product OR services
- some posts are plain spams
Coincidence or Fake?

Reviews for a moving company from YellowPages

- # of merchants reviewed by each of these reviewers ➔ 1
- Review dates close to one another
- All rated 5 star
- Reviewers seem to know exact names of people working in the company and TOO many positive mentions
So where does all of this lead to?
Heard of these terms?

Subjectivity Analysis

Review Mining

Opinion Mining

Sentiment Analysis

Appraisal Extraction

Synonymous & Interchangeably Used!
So, what is Subjectivity?

• The **linguistic** expression of somebody’s **opinions**, **sentiments**, **emotions**…..(private states)

• **private state**: state that is not open to objective verification  
  (Quirk, Greenbaum, Leech, Svartvik (1985). A Comprehensive Grammar of the English Language.)

• **Subjectivity analysis** - is the computational study of **affect**, **opinions**, and **sentiments** expressed in text
  
  • blogs
  • editorials
  • reviews (of products, movies, books, etc.)
  • newspaper articles
Example: iPhone review

Review on InfoWorld - tech news site

- summary is structured
- everything else is plain text
- mixture of objective and subjective information
- no separation between positives and negatives

CNET review

- nice structure
- positives and negatives separated

Tech BLOG

- everything is plain text
- no separation between positives and negatives
Lab test: Apple gets iPhone 3G right for business
An abundance of new features carries iPhone 3G and iPhone 2.0 into the enterprise

By Tom Yager
July 24, 2008

With the iPhone 3G’s banner opening weekend and newsstands looking like a rack of brochures for the device, a review of the iPhone 3G at this point might be pro forma, except for one thing: Much of the iPhone 3G and the new iPhone 2.0 software remains an enigma to professionals and enterprises, users set apart by, among other things, their tendency to use punctuation in their e-mail. These users demand more from a handset than a cellular browser and YouTube.

Related Stories
New MacBook Air: now with extra SSD goodness
AT&T says Phone 3G tethering coming ‘soon’
Popular Tags
apple, iphone-3g

See Also
iPhone delivers more misses than hits
iPhone: The $1,375 Pod
Back to special report: Apple launches the iPhone 3G

The Bottom Line
Apple iPhone 3G
Apple, apple.com/iphone

Very Good 8.5
Criteria score weight
Extensibility 7 20%
Messaging 8 20%
Networking 9 20%
Usability 9 20%
Multimedia 9 20%
Value 8 20%

Product summary

The good:
The Apple iPhone has a stunning display, a sleek design, and an innovative multitouch user interface. Its Safari browser makes for a superb Web surfing experience, and it offers easy-to-use apps. As an iPod, it shines.

The bad:
The Apple iPhone has variable call quality and lacks some basic features found in many cell phones, including stereo Bluetooth support and 3G compatibility. Integrated memory is stingy for an iPod, and you have to sync the iPhone to manage music content.

The bottom line:
Despite some important missing features, a slow data network, and call quality that doesn’t always deliver, the Apple iPhone sets a new benchmark for an integrated cell phone and MP3 player.

Specifications:
OS provided: Apple Mac OS X; Band / mode: GSM 850/900/1800/1900 (Quadband); Wireless connectivity: IEEE 802.11b, IEEE 802.11g, Bluetooth 2.0 EDR; See full specs
See all products in the Apple iPhone series

CNET editors’ review
Reviewed by: Kent Gerner
Edited by: Lindsey Turner
Reviewed on: 08/30/2007
Updated on: 07/11/2008

Review on InfoWorld - tech news site

Example: iPhone review

CNET review

Review posted on a tech blog

See my NEW iPhone 3G review.

Let me start off by saying that while I’m a fan of Apple’s success and products, I’m not one of those people that blindly apologizes for their products no matter what. I’ll be the first to say that something works or it doesn’t. My friends and many of you come to me all the time because they want my HONEST assessment. So I wanted a couple of days with the iPhone to really take it through its paces and see if this new phone is what it’s hyped up to be. You must also understand that there isn’t a smartphone out there that I think is perfect. As a matter of fact before the iPhone there were basically 4 smartphone OS’s, Palm, BlackBerry, Symbian and Windows Mobile. I stuck with Palm because it was the lessor of the 4 evils or the one that sucks least. Palm has a UI (user interface) that is actually zero innovation. However, there are thousands of apps for it. BlackBerry doesn’t have a touch screen or tap pad/thumb wheel. Also the Blackberry’s 1 considered case is the Bold (etc.) Symbian looked very promising, but I was always afraid how slooooww it was and that there were very few Symbians around. So when I got the phone I was impressed when it showed up on him just last week right in front of me.
Subjectivity Analysis on iPhone Reviews

Individual’s Perspective

• Highlight of what is good and bad about iPhone
  • Ex. Tech blog may contain mixture of information
• Combination of good and bad from the different sites *(tech blog, InfoWorld and CNET)*
  • Complementing information
  • Contrasting opinions
  Ex.

  **CNET**: *The iPhone lacks some basic features*

  **Tech Blog**: *The iPhone has a complete set of features*
Subjectivity Analysis on iPhone Reviews

Business’ Perspective
- **Apple**: What do consumers think about iPhone?
  - Do they like it?
  - What do they dislike?
  - What are the major complaints?
  - What features should we add?
- **Apple’s competitor**:
  - What are iPhone’s weaknesses?
  - How can we compete with them?
  - Do people like everything about it?
Opinion Trend (temporal)  
Sentiments for a given product/brand/services

Free 14 day trial - Sign Up

Just some of the great benefits:

- Monitor what is being said about you, your brand and your products, or services across blogs, forums, discussion boards and news sites
- Discover the sentiment and opinions of these people
- View reports on gender, age groups and location
- View the main influencers of your brand
- Compare how you are doing to your competitors.
- Setup alerts to email you when bad press events occur
- Gain intelligence to develop more targeted marketing, improve products and gain even more customers!

Click the links below to view just some of the reports you will have immediate access to:

Summary Buzz Volume Sentiment Topics Influencers Age Gender Location Media

Free Trial Register Today!

Simply complete your details and you can be using our state of the art system in a few minutes, absolutely free for 14 days.

Email Address
(Account details sent here)

Password

Security Image:

By completing this information you are agreeing to our terms and conditions

Register >>

Are you from an agency? Click here to book a demo instead

Business Intelligence Software
Application Areas Summarized

- Businesses and organizations: interested in opinions
  - product and service benchmarking
  - market intelligence
  - survey on a topic
- Individuals: interested in other’s opinions when
  - Purchasing a product
  - Using a service
  - Tracking political topics
  - Other decision making tasks
- Ads placements: Placing ads in user-generated content
  - Place an ad when one praises an product
  - Place an ad from a competitor if one criticizes a product
- Opinion search: providing general search for opinions
OUTLINE

• Emotion Detection
• Subjectivity Overview
• Sentiment Analysis
• Opinion Mining
SENTIMENT ANALYSIS

- Definition
- Annotation
- Lexical Resources
- Supervised Models
- Unsupervised Models
- Social Media
FLAVORS OF SUBJECTIVITY
ANALYSIS

Synonyms and Used Interchangeably !!

Sentiment Analysis
Opinion Mining
Mood Classification
Emotion Analysis
**BASICS ..**

- Basic components
  - Opinion Holder – Who is talking?
  - Object – Item on which opinion is expressed.
  - Opinion – Attitude or view of the opinion holder.

“This is a good book.”

Opinion

Opinion Holder

Object
Review Websites

- www.burrrp.com
- www.mouthshut.com
- www.justdial.com
- www.yelp.com
- www.zagat.com
- www.bollywoodhungama.com
- www.indya.com

Restaurant reviews (now, for a variety of ‘lifestyle’ products/services)

A wide variety of reviews

Prof. reviews: Well-formed
User reviews: More mistakes

Movie reviews by professional critics, users. Links to external reviews also present
A typical Review website

[Image of a review website with ratings, comments, and details about a university review.

Pros: It's a good experience.
Cons: It's not a suggestion to be in IITB only.

MouthShut Product Rating: Recommended by 80% members

Member's Rating: ★★★★★
Member's Recommendation: No
Read 802 times
Rated by 5 members

Write your own review

IIT Bombay, this name makes my blood boil - forces in the positive sense. I spent my fabulous 4 years of life in there and I assure one and all of you this is the place to be in.

Let's start with how it feels to be in there. For this I would like to describe my first week in there.

This was the first time I was in Bombay... date: 16th July 2001. My heart was beating like anything when I reached the main gate and saw the logo on the gate with the motto Gyaanam Param Dhayaam. I was enticed by the very first look...]
Sample Review 1
(This, that and this)

• FLY E300 is a good mobile which I purchased recently with lots of hesitation. Since this brand is not familiar in the market as well known as Sony Ericsson, the features for a good mobile. Any other brand almost all 19k Indian rupees. But this one is only 9k.

Touch Screen, good resolution, good talk time, 3.2Mega Pixel camera, A2DP, IRDA and so on...

BUT BEWARE THAT THE CAMERA IS NOT THAT GOOD, THOUGH IT FEATURES 3.2 MEGA PIXEL, IT'S NOT AS GOOD AS MY PREVIOUS MOBILE SONY ERICSSON K750i which is just 2Mega Pixel.

Sony Ericsson was excellent with the feature of camera. So if anyone is thinking for camera, please excuse. This model of FLY is not apt for you. Am fooled in this regard.

Audio is not bad, infact better than Sony Ericsson K750i.

FLY is not user friendly probably since we have just started to use this brand.

‘Touch screen’ today signifies a positive feature. Will it be the same in the future?

Comparing old products

The confused conclusion

From: www.mouthshut.com
Sample Review 2
(Noise)

Hi,

I have Haier phone.. It was good when i was buing this phone.. But I invented A lot of bad features by this phone those are It’s cost is low but Software is not good and Battery is very bad,,, Ther are no signals at out side of the city,,, People can’t understand this type of software,,, There aren’t features in this phone, Design is better not good,,, Sound also bad.. So I’m not intrest this side They are giving heare phones it is good. They are giving more talktime and validity these are also good. They are giving colour screen at display time it is also good because other phones aren’t these features. So I am still wait.

From: www.mouthshut.com
Sample Review 3
(Alternating sentiments)

I suggest that instead of fillings songs in tunes you should fill tunes (not made of songs) only. The phone has good popularity in old age people. Third i had tried much of its data cable but i find it nowhere. It should be supplied with set with some extra cost.

Good features of this phone are its cheapest price and durability. It should have some features more than nokia 1200. It is easily available in market and repair is also available.

From: www.mouthshut.com
Sample Review 4
(Subject-centric or not?)

- I have this personal experience of using this cell phone. I bought it one and half years back. It had modern features that a normal cell phone has, and the look is excellent. I was very impressed by the design. I bought it for Rs. 8000. It was a gift for someone. It worked fine for first one month, and then started the series of multiple faults it has. First the speaker didnt work, I took it to the service centre (which is like a govt. office with no work). It took 15 days to repair the handset, moreover they charged me Rs. 500. Then after 15 days again the mike didnt work, then again same set of time was consumed for the repairs and it continued. Later the camera didnt work, the speaks were rubbish, it used to hang. It started restarting automatically. And the govt. office had staff which I doubt have any knoledge of cell phones??

These multiple faults continued for as long as one year, when the warranty period ended. In this period of time I spent a considerable amount on the petrol, a lot of time (as the service centre is a govt. office). And at last the phone is still working, but now it works as a paper weight. The company who produces such items must be sacked. I understand that it might be fault with one prticular handset, but the company itself never bothered for replacement and I have never seen such miserable cust service. For a comman man like me, Rs. 8000 is a big amount. And I spent almost the same amount to get it work, if any has a good suggestion and can gude me how to sue such companies, please guide.

For this the quality team is faulty, the cust service is really miserable and the worst condition of any organisation I have ever seen is with the service centre for Fly and Sony Erricson, (it's near Sancheti hospital, Pune). I dont have any thing else to say.

From: www.mouthshut.com
Sample Review 5
(Good old sarcasm)

“I’ve seen movies where there was practically no plot besides explosion, explosion, catchphrase, explosion. I’ve even seen a movie where nothing happens. But *White on Rice* was new on me: a collection of really wonderful and appealing characters doing completely baffling and uncharacteristic things.”

Review from: www.pajiba.com
Q: What is the international reaction to the reelection of Robert Mugabe as President of Zimbabwe?

A: African observers generally approved of his victory while Western Governments denounced it.
More motivations

- **Product review mining**: What features of the ThinkPad T43 do customers like and which do they dislike?
- **Review classification**: Is a review positive or negative toward the movie?
- **Tracking sentiments toward topics over time**: Is anger ratcheting up or cooling down?
- **Etc.**
“The report is full of absurdities,” Xirao-Nima said the next day.
“The report is full of absurdities,” Xirao-Nima said the next day.

**Objective speech event**
- anchor: the entire sentence
- source: <writer>
- implicit: true

**Direct subjective**
- anchor: said
- source: <writer, Xirao-Nima>
- intensity: high
- expression intensity: neutral
- attitude type: negative
- target: report

**Expressive subjective element**
- anchor: full of absurdities
- source: <writer, Xirao-Nima>
- intensity: high
- attitude type: negative
“The report is full of absurdities,” Xirao-Nima said the next day.

**Objective speech event**
- anchor: the entire sentence
- source: <writer>
- implicit: true

**Direct subjective**
- anchor: said
- source: <writer, Xirao-Nima>
- intensity: high
- expression intensity: neutral
- attitude type: negative
- target: report

**Expressive subjective element**
- anchor: full of absurdities
- source: <writer, Xirao-Nima>
- intensity: high
- attitude type: negative
“The report is **full of absurdities,**” Xirao-Nima said the next day.
TYPES OF OPINIONS

• Direct
  • “This is a great book.”
  • “Mobile with awesome functions.”

• Comparison
  • “Samsung Galaxy S3 is better than Apple iPhone 4S.”
  • “Hyundai Eon is not as good as Maruti Alto ! .”
WHAT IS SENTIMENT CLASSIFICATION

• Classify given text on the overall sentiments expresses by the author
• Different levels
  • Document
  • Sentence
  • Feature
• Classification levels
  • Binary
  • Multi Class
DOCUMENT LEVEL SENTIMENT CLASSIFICATION

- Documents can be reviews, blog posts,..
- Assumption:
  - Each document focuses on single object.
  - Only single opinion holder.
- Task: determine the overall sentiment orientation of the document.
SENTENCE LEVEL SENTIMENT CLASSIFICATION

- Considers each sentence as a separate unit.
- Assumption: sentence contain only one opinion.
- Task 1: identify if sentence is subjective or objective
- Task 2: identify polarity of sentence.
FEATURE LEVEL SENTIMENT CLASSIFICATION

- Task 1: identify and extract object features
- Task 2: determine polarity of opinions on features
- Task 3: group same features
- Task 4: summarization

- Ex. This mobile has *good* camera but *poor* battery life.
APPROACHES

• Prior Learning
• Subjective Lexicon
• (Un)Supervised Machine Learning
APPROACH 1: PRIOR LEARNING

• Utilize available pre-annotated data
  • Amazon Product Review (star rated)
  • Twitter Dataset(s)
  • IMDb movie reviews (star rated)
• Learn keywords, N-Gram with polarity
KEYWORDS SELECTION FROM TEXT

- Pang et. al. (2002)
  - Two human’s hired to pick keywords
  - Binary Classification of Keywords
    - Positive
    - Negative
  - Unigram method reached 80% accuracy.
N-GRAM BASED CLASSIFICATION

• Learn N-Grams (frequencies) from pre-annotated training data.
• Use this model to classify new incoming sample.
• Classification can be done using
  • Counting method
  • Scoring function(s)
PART-OF-SPEECH BASED PATTERNS

- Extract POS patterns from training data.
- Usually used for subjective vs objective classification.
- Adjectives and Adverbs contain sentiments.
- Example patterns
  - *-JJ-NN : trigram pattern
  - JJ-NNP : bigram pattern
  - *-JJ : bigram pattern
SUBJECTIVE LEXICON

- Heuristic or Hand Made
- Can be General or Domain Specific
- Difficult to Create
- Sample Lexicons
  - General Inquirer (1966)
  - Dictionary of Affective Language
  - SentiWordNet (2006)
GENERAL INQUIRER

- Positive and Negative connotations.
- List of words manually created.
  - 1915 Positive Words
  - 2291 Negative Words
- http://wjh.harvard.edu/~inquirer
DICTIONARY OF AFFECTIVE LANGUAGE

- 9000 Words with Part-of-speech information
- Each word has a valance score range 1 – 3.
  - 1 for Negative
  - 3 for Positive
- App
- [http://sail.usc.edu/~kazemzad/emotion_in_text/cgi/DAL_app/index.php](http://sail.usc.edu/~kazemzad/emotion_in_text/cgi/DAL_app/index.php)
SENTIWORDNET

- Approx 1.7 Million words
- Using WordNet and Ternary Classifier.
- Classifier is based on Bag-of-Synset model.
- Each synset is assigned three scores
  - Positive
  - Negative
  - Objective
EXAMPLE : SCORES FROM SENTIWORDNET

• Very comfortable, but straps go loose quickly.
  • comfortable
    • Positive: 0.75
    • Objective: 0.25
    • Negative: 0.0
  • loose
    • Positive: 0.0
    • Objective: 0.375
    • Negative: 0.625

• Overall - Positive
  • Positive: 0.75
  • Objective: 0.625
  • Negative: 0.625
ADVANTAGES AND DISADVANTAGES

• Advantages
  • Fast
  • No Training data necessary
  • Good initial accuracy

• Disadvantages
  • Does not deal with multiple word senses
  • Does not work for multiple word phrases
MACHINE LEARNING

- Sensitive to sparse and insufficient data.
- Supervised methods require annotated data.
- Training data is used to create a hyper plane between the two classes.
- New instances are classified by finding their position on hyper plane.
MACHINE LEARNING

- SVMs are widely used ML Technique for creating feature-vector-based classifiers.
- Commonly used features
  - N-Grams or Keywords
    - Presence: Binary
    - Count: Real Numbers
  - Special Symbols like !, ?, @, #, etc.
  - Smiley
SOME UNANSWERED QUESTIONS!

- Sarcasm Handling
- Word Sense Disambiguation
- Pre-processing and cleaning
- Multi-class classification
CHALLENGES

• Negation Handling
  • I don’t like Apple products.
  • This is not a good read.
• Un-Structured Data, Slangs, Abbreviations
  • Lol, rofl, omg! …..
  • Gr8, IMHO, …
• Noise
  • Smiley
  • Special Symbols ( ! , ? , .... )
CHALLENGES

• Ambiguous words
  • This music cd is literal waste of time. (negative)
  • Please throw your waste material here. (neutral)
• Sarcasm detection and handling
  • “All the features you want - too bad they don’t work. :-P”
• (Almost) No resources and tools for low/scarce resource languages like Indian languages.
DATASETS

• Movie Review Dataset
  • Bo Pang and Lillian Lee
  • http://www.cs.cornell.edu/People/pabo/movie-review-data/

• Product Review Dataset
  • Blitzer et. al.
  • Amazon.com product reviews
  • 25 product domains
  • http://www.cs.jhu.edu/~mdredze/datasets/sentiment
DATASETS

• MPQA Corpus
  • Multi Perspective Question Answering
  • News Article, other text documents
  • Manually annotated
  • 692 documents
• Twitter Dataset
  • http://www.sentiment140.com/
  • 1.6 million annotated tweets
  • Bi-Polar classification
Corpus

- [www.cs.pitt.edu/mqpa/databaserelease](http://www.cs.pitt.edu/mqpa/databaserelease) (version 2)

- English language versions of articles from the world press (187 news sources)

- Also includes contextual polarity annotations (later)

- Themes of the instructions
  - No rules about how particular words should be annotated.
  
  - Don’t take expressions out of context and think about what they could mean, but judge them as they are used in that sentence.
Who does lexicon development?

- Humans
- Semi-automatic
- Fully automatic
What?

- Find relevant words, phrases, patterns that can be used to express subjectivity
- Determine the polarity of subjective expressions
Words

  - positive: honest important mature large patient
  
  - Ron Paul is the only **honest** man in Washington.
  - Kitchell’s writing is unbelievably **mature** and is only likely to get better.
  - To humour me my **patient** father agrees yet again to my choice of film
Words

  - positive
  - negative: harmful hypocritical inefficient insecure
    - It was a macabre and hypocritical circus.
    - Why are they being so inefficient?
Words

  - positive
  - negative
  - Subjective (but not positive or negative sentiment): curious, peculiar, odd, likely, probable
    - He spoke of Sue as his probable successor.
    - The two species are likely to flower at different times.
• **Other parts of speech** (e.g. Turney & Littman 2003, Riloff, Wiebe & Wilson 2003, Esuli & Sebastiani 2006)
  
  • **Verbs**
    • positive: *praise, love*
    • negative: *blame, criticize*
    • subjective: *predict*
  
  • **Nouns**
    • positive: *pleasure, enjoyment*
    • negative: *pain, criticism*
    • subjective: *prediction, feeling*
Phrases

• Phrases containing adjectives and adverbs (e.g. Turney 2002, Takamura, Inui & Okumura 2007)
  • positive: high intelligence, low cost
  • negative: little variation, many troubles
Patterns

- **Lexico-syntactic patterns**  (Riloff & Wiebe 2003)
- **way with <np>**: … to ever let China use force to have its way with …
- **expense of <np>**: at the expense of the world’s security and stability
- **underlined <dobj>**: Jiang’s subdued tone … underlined his desire to avoid disputes …
How?

- How do we identify subjective items?
How?

- How do we identify subjective items?

- Assume that contexts are coherent
Conjunction

The Homestay Experience - Cultural Kaleidoscope 2006
My host's home was very nice and comfortable. I got to try all types of food; Malaysian, Chinese, Indonesian and I loved it all. My host's parents were very...
www.gardenschool.edu.my/studentportal/aec/Kaleidoscope06/experience.asp - 10k - Cached - Similar pages - Note this

PriceGrabber User Rating for Watch Your Budget - PriceGrabber.com
Reviews, Camera I purchased was very nice and a bargain. There was a problem with shipping, but was resolved quickly. Buy with confidence from this vendor...
www.pricegrabber.com/rating_getreview.php?retid=5821 - Similar pages - Note this

Testimonials
"Everybody was very nice and service was as fast as they possibly could... "Staff member who helped me was very nice and easy to talk to."...
www.sa.psu.edu/uhs/news/testimonials.cfm - 22k - Cached - Similar pages - Note this

Naxos Villages - Naxos Town or Chora Reviews: Very nice and very...
-Did you enjoy the trip to Naxos Town: Yes it was very nice and very scenic. In order to get to the village were there enough signs in order to find it: It...
Statistical association

• If words of the same orientation like to co-occur together, then the presence of one makes the other more probable.

• Use statistical measures of association to capture this interdependence.
  • E.g., Mutual Information (Church & Hanks 1989)
How?

• How do we identify subjective items?

• Assume that contexts are coherent
• Assume that alternatives are similarly subjective
How?

- How do we identify subjective items?
- Assume that contexts are coherent
- Assume that alternatives are similarly subjective
WordNet

(7) S: (adj) brainy, brilliant, smart as a whip (having or marked by unusual and impressive intelligence) "some men dislike brainy women"; "a brilliant mind"; "a brilliant solution to the problem"

  similar to
  S: (adj) intelligent (having the capacity for thought and reason especially to a high degree) "is there intelligent life in the universe?"; "an intelligent question"

  derivationally related form
  W: (n) brilliancy [Related to: brilliant] (a quality that outshines the usual)
  W: (n) brilliance [Related to: brilliant] (unusual mental ability)

  antonym
  W: (adj) unintelligent [Indirect via intelligent] (lacking intelligence) "a dull job with lazy and unintelligent co-workers"
WordNet

♦ (7) **S:** (adj) 
  - brainy, brilliant, smart as a whip (having or marked by unusual and impressive intelligence) “some men dislike brainy women”, “a brilliant mind”; “a brilliant solution to the problem”
    - similar to
      ♦ **S:** (adj) intelligent (having the capacity for thought and reason especially to a high degree) “is there intelligent life in the universe?”; “an intelligent question”
    - derivationally related form
      ♦ **W:** (n) brilliancy [Related to: brilliant] (a quality that outshines the usual)
      ♦ **W:** (n) brilliance [Related to: brilliant] (unusual mental ability)
  - antonym
    ♦ **W:** (adj) unintelligent [Indirect via intelligent] (lacking intelligence) “a dull job with lazy and unintelligent co-workers”
WordNet relations

(7) S: (adj) brainy, brilliant, smart as a whip (having or marked by unusual and impressive intelligence) "some men dislike brainy women"; "a brilliant mind"; "a brilliant solution to the problem"
  - similar to
    - S: (adj) intelligent (having the capacity for thought and reason especially to a high degree) "is there intelligent life in the universe?"; "an intelligent question"
  - derivationally related form
    - W: (n) brilliancy [Related to: brilliant] (a quality that outshines the usual)
    - W: (n) brilliance [Related to: brilliant] (unusual mental ability)
  - antonym
    - W: (adj) unintelligent [Indirect via intelligent] (lacking intelligence) "a dull job with lazy and unintelligent co-workers"
WordNet relations

(7) S: (adj) brainy, brilliant, smart as a whip (having or marked by unusual and impressive intelligence) "some men dislike brainy women"; "a brilliant mind"; "a brilliant solution to the problem"
  - similar to
    - S: (adj) intelligent (having the capacity for thought and reason especially to a high degree) "is there intelligent life in the universe?"; "an intelligent question"
  - derivationally related form
    - W: (n) brilliancy [Related to: brilliant] (a quality that outshines the usual)
    - W: (n) brilliance [Related to: brilliant] (unusual mental ability)
  - antonym
    - W: (adj) unintelligent [Indirect via intelligent] (lacking intelligence) "a dull job with lazy and unintelligent co-workers"
WordNet relations

(7) S: (adj) brainy, brilliant, smart as a whip (having or marked by unusual and impressive intelligence) "some men dislike brainy women"; "a brilliant mind"; "a brilliant solution to the problem"
  - similar to
    - S: (adj) intelligent (having the capacity for thought and reason especially to a high degree) "is there intelligent life in the universe?"; "an intelligent question"
  - derivationally related form
    - W: (n) brilliancy [Related to: brilliant] (a quality that outshines the usual)
    - W: (n) brilliance [Related to: brilliant] (unusual mental ability)
  - antonym
    - W: (adj) unintelligent [Indirect via intelligent] (lacking intelligence) "a dull job with lazy and unintelligent co-workers"
WordNet glosses

(7) S: (adj) brainy, brilliant, smart as a whip (having or marked by unusual and impressive intelligence) "some men dislike brainy women"; "a brilliant mind", "a brilliant solution to the problem"

- similar to
  - S: (adj) intelligent (having the capacity for thought and reason especially to a high degree) "is there intelligent life in the universe?"; "an intelligent question"

- derivationally related form
  - W: (n) brilliancy [Related to: brilliant] (a quality that outshines the usual)
  - W: (n) brilliance [Related to: brilliant] (unusual mental ability)

- antonym
  - W: (adj) unintelligent [Indirect via intelligent] (lacking intelligence) "a dull job with lazy and unintelligent co-workers"
WordNet examples

(7) S: (adj) **brainless, brilliant, smart as a whip** (having or marked by unusual and impressive intelligence) "some men dislike brainy women"; "a brilliant mind"; "a brilliant solution to the problem"

- **similar to**
  - S: (adj) **intelligent** (having the capacity for thought and reason especially to a high degree) "is there intelligent life in the universe?"; "an intelligent question"

- **derivationally related form**
  - W: (n) **brilliancy** [Related to: brilliant] (a quality that outshines the usual)
  - W: (n) **brilliance** [Related to: brilliant] (unusual mental ability)

- **antonym**
  - W: (adj) **unintelligent** [Indirect via intelligent] (lacking intelligence) "a dull job with lazy and unintelligent co-workers"
Extracting Opinions

Which words are opinion words?

PROXIMITY

• Opinion words are adjectives and adverbs.
• Likely to be opinions if they occur near a feature mention.
• Computationally cheap.
• Negation is hard to detect.
• Imprecise.

DEPENDENCY PARSING

• Opinion words are adjectives and adverbs.
• Likely to be opinions if amod / nsubj / advmod relationship exists to feature mention.
• Computationally expensive.
• neg (negation) relations are easily detected.
• Precise.
Extracting Opinions

"The controls are intuitive."

There are large controls on the top.

"The controls feel natural."

How to classify adjectives?
Scoring Product Features

explicit
product
feature

referring
opinions

product feature score
Classifying Opinions

- Synonymous words have high Web-PMI with each other

\[
\text{WebPMI(adj, great)} = \frac{\text{HITS(“camera” near adj, great)}}{\text{HITS(“camera” NEAR adj) x HITS(“camera” NEAR great)}}
\]

F1 Scores: 0.78(+) 0.76(-)
Accounting for Negation

• Let us consider the following positive sentence:
  • Example: *Luckily, the smelly poo did not leave awfully nasty stains* on my favorite shoes!

• Rest of Sentence (RoS):
  • Following: *Luckily, the smelly poo did not leave awfully nasty stains* on my favorite shoes!
  • Around: *Luckily, the smelly poo did not leave awfully nasty stains* on my favorite shoes!

• First Sentiment-Carrying Word (FSW):
  • Following: *Luckily, the smelly poo did not leave awfully nasty stains* on my favorite shoes!
  • Around: *Luckily, the smelly poo did not leave awfully nasty stains* on my favorite shoes!

SMC 2011
Accounting for Negation

• Let us consider the following positive sentence:
  • Example: *Luckily, the smelly poo did not leave awfully nasty stains on my favorite shoes!*

• Next Non-Adverb (NNA):
  • Following: *Luckily, the smelly poo did not leave awfully nasty stains on my favorite shoes!*

• Fixed Window Length (FWL):
  • Following (3): *Luckily, the smelly poo did not leave awfully nasty stains on my favorite shoes!*
  • Around (3): *Luckily, the smelly poo did not leave awfully nasty stains on my favorite shoes!*

SMC 2011
Framework (1)

- Lexicon-based sentence-level sentiment scoring by using SentiWordNet
- Optional support for sentiment negation
- Individual words are scored in the range [-1,1]
- Word scores are used to classify a sentence as positive (1) or negative (-1)
Framework (2)

- Score sentences in test corpus for their sentiment
- For an arbitrary sentence:
  - Retrieve all words (simple and compound)
  - Retrieve each words’ Part-Of-Speech (POS) and lemma
  - Disambiguate word senses (Lesk algorithm)
  - Retrieve words’ sentiment scores from lexicon
  - Negate sentiment scores of negated words, as determined by means of one of the considered approaches, by multiplying the scores with an inversion factor (typically negative)
  - Calculate sentence score as sum of words’ scores
  - Classify sentence as either positive (score $\geq 0$) or negative (score $< 0$)
How? Summary

• How do we identify subjective items?

• Assume that contexts are coherent
• Assume that alternatives are similarly subjective
• Take advantage of word meanings
<table>
<thead>
<tr>
<th>object</th>
<th>15651</th>
<th>5.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>damage</td>
<td>938</td>
<td>10.09</td>
</tr>
<tr>
<td>harm</td>
<td>276</td>
<td>8.91</td>
</tr>
<tr>
<td>injury</td>
<td>295</td>
<td>8.38</td>
</tr>
<tr>
<td>problem</td>
<td>1014</td>
<td>8.37</td>
</tr>
<tr>
<td>trouble</td>
<td>249</td>
<td>8.32</td>
</tr>
<tr>
<td>death</td>
<td>383</td>
<td>7.96</td>
</tr>
<tr>
<td>delay</td>
<td>146</td>
<td>7.87</td>
</tr>
<tr>
<td>confusion</td>
<td>137</td>
<td>7.8</td>
</tr>
<tr>
<td>difficulty</td>
<td>223</td>
<td>7.74</td>
</tr>
<tr>
<td>disruption</td>
<td>111</td>
<td>7.71</td>
</tr>
<tr>
<td>distress</td>
<td>101</td>
<td>7.52</td>
</tr>
<tr>
<td>concern</td>
<td>190</td>
<td>7.35</td>
</tr>
<tr>
<td>pain</td>
<td>126</td>
<td>7.24</td>
</tr>
<tr>
<td>chaos</td>
<td>82</td>
<td>7.22</td>
</tr>
<tr>
<td>accident</td>
<td>126</td>
<td>7.22</td>
</tr>
<tr>
<td>loss</td>
<td>190</td>
<td>7.14</td>
</tr>
<tr>
<td>controversy</td>
<td>81</td>
<td>7.12</td>
</tr>
<tr>
<td>pollution</td>
<td>88</td>
<td>7.04</td>
</tr>
<tr>
<td>havoc</td>
<td>64</td>
<td>7.01</td>
</tr>
<tr>
<td>cancer</td>
<td>93</td>
<td>7.01</td>
</tr>
<tr>
<td>stir</td>
<td>62</td>
<td>7.0</td>
</tr>
<tr>
<td>suffering</td>
<td>70</td>
<td>6.99</td>
</tr>
<tr>
<td>disease</td>
<td>141</td>
<td>6.95</td>
</tr>
<tr>
<td>explosion</td>
<td>72</td>
<td>6.93</td>
</tr>
<tr>
<td>embarrassment</td>
<td>63</td>
<td>6.87</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>object</th>
<th>9100</th>
<th>6.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>negligence</td>
<td>55</td>
<td>7.34</td>
</tr>
<tr>
<td>virus</td>
<td>53</td>
<td>7.14</td>
</tr>
<tr>
<td>smoking</td>
<td>27</td>
<td>6.36</td>
</tr>
<tr>
<td>defect</td>
<td>29</td>
<td>6.32</td>
</tr>
<tr>
<td>bacteria</td>
<td>26</td>
<td>6.23</td>
</tr>
<tr>
<td>infection</td>
<td>32</td>
<td>6.09</td>
</tr>
<tr>
<td>factor</td>
<td>76</td>
<td>6.07</td>
</tr>
<tr>
<td>assault</td>
<td>28</td>
<td>6.05</td>
</tr>
<tr>
<td>pollution</td>
<td>31</td>
<td>6.04</td>
</tr>
<tr>
<td>recession</td>
<td>28</td>
<td>5.99</td>
</tr>
<tr>
<td>stress</td>
<td>28</td>
<td>5.88</td>
</tr>
<tr>
<td>accident</td>
<td>36</td>
<td>5.8</td>
</tr>
<tr>
<td>bomb</td>
<td>26</td>
<td>5.78</td>
</tr>
<tr>
<td>disease</td>
<td>50</td>
<td>5.74</td>
</tr>
<tr>
<td>fire</td>
<td>45</td>
<td>5.63</td>
</tr>
<tr>
<td>lack</td>
<td>37</td>
<td>5.6</td>
</tr>
<tr>
<td>organism</td>
<td>18</td>
<td>5.58</td>
</tr>
<tr>
<td>deficiency</td>
<td>16</td>
<td>5.57</td>
</tr>
<tr>
<td>fault</td>
<td>21</td>
<td>5.57</td>
</tr>
<tr>
<td>delay</td>
<td>20</td>
<td>5.55</td>
</tr>
<tr>
<td>damage</td>
<td>30</td>
<td>5.51</td>
</tr>
<tr>
<td>weather</td>
<td>22</td>
<td>5.39</td>
</tr>
<tr>
<td>explosion</td>
<td>15</td>
<td>5.27</td>
</tr>
<tr>
<td>drought</td>
<td>12</td>
<td>5.27</td>
</tr>
<tr>
<td>parasite</td>
<td>12</td>
<td>5.26</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>modifier</th>
<th>1971</th>
<th>1.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>reasonable</td>
<td>26</td>
<td>8.72</td>
</tr>
<tr>
<td>indirectly</td>
<td>16</td>
<td>7.77</td>
</tr>
<tr>
<td>possibly</td>
<td>27</td>
<td>7.67</td>
</tr>
<tr>
<td>thereby</td>
<td>26</td>
<td>7.66</td>
</tr>
<tr>
<td>mainly</td>
<td>32</td>
<td>7.51</td>
</tr>
<tr>
<td>inevitably</td>
<td>20</td>
<td>7.48</td>
</tr>
<tr>
<td>permit</td>
<td>18</td>
<td>6.09</td>
</tr>
<tr>
<td>contribute</td>
<td>5</td>
<td>4.21</td>
</tr>
<tr>
<td>use</td>
<td>6</td>
<td>0.22</td>
</tr>
<tr>
<td>part</td>
<td>10</td>
<td>0.0</td>
</tr>
<tr>
<td>np_VPto</td>
<td>2407</td>
<td>25.0</td>
</tr>
<tr>
<td>prep_Sing</td>
<td>158</td>
<td>11.3</td>
</tr>
<tr>
<td>np_pp</td>
<td>4585</td>
<td>5.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>and/or</th>
<th>130</th>
<th>0.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>permit</td>
<td>18</td>
<td>6.09</td>
</tr>
<tr>
<td>contribute</td>
<td>5</td>
<td>4.21</td>
</tr>
<tr>
<td>use</td>
<td>6</td>
<td>0.22</td>
</tr>
<tr>
<td>np_VPto</td>
<td>2407</td>
<td>25.0</td>
</tr>
<tr>
<td>prep_Sing</td>
<td>158</td>
<td>11.3</td>
</tr>
<tr>
<td>np_pp</td>
<td>4585</td>
<td>5.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>pp</th>
<th>3374</th>
<th>16.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>negligence</td>
<td>46</td>
<td>8.14</td>
</tr>
<tr>
<td>defect</td>
<td>21</td>
<td>6.81</td>
</tr>
<tr>
<td>bacterium</td>
<td>17</td>
<td>6.62</td>
</tr>
<tr>
<td>virus</td>
<td>17</td>
<td>6.4</td>
</tr>
<tr>
<td>smoking</td>
<td>13</td>
<td>6.4</td>
</tr>
<tr>
<td>fault</td>
<td>19</td>
<td>6.15</td>
</tr>
<tr>
<td>lack</td>
<td>37</td>
<td>5.98</td>
</tr>
<tr>
<td>deficiency</td>
<td>10</td>
<td>5.95</td>
</tr>
<tr>
<td>shortage</td>
<td>12</td>
<td>5.75</td>
</tr>
<tr>
<td>blockage</td>
<td>6</td>
<td>5.71</td>
</tr>
<tr>
<td>breach</td>
<td>14</td>
<td>5.66</td>
</tr>
<tr>
<td>parasite</td>
<td>7</td>
<td>5.65</td>
</tr>
<tr>
<td>default</td>
<td>7</td>
<td>5.59</td>
</tr>
<tr>
<td>error</td>
<td>18</td>
<td>5.58</td>
</tr>
<tr>
<td>abnormality</td>
<td>7</td>
<td>5.58</td>
</tr>
<tr>
<td>theft</td>
<td>9</td>
<td>5.58</td>
</tr>
<tr>
<td>pollution</td>
<td>14</td>
<td>5.57</td>
</tr>
<tr>
<td>enteritis</td>
<td>5</td>
<td>5.53</td>
</tr>
<tr>
<td>build-up</td>
<td>6</td>
<td>5.51</td>
</tr>
<tr>
<td>fall</td>
<td>17</td>
<td>5.55</td>
</tr>
<tr>
<td>exposure</td>
<td>11</td>
<td>5.5</td>
</tr>
<tr>
<td>warming</td>
<td>6</td>
<td>5.46</td>
</tr>
<tr>
<td>drought</td>
<td>6</td>
<td>5.45</td>
</tr>
<tr>
<td>failure</td>
<td>24</td>
<td>5.38</td>
</tr>
<tr>
<td>recession</td>
<td>11</td>
<td>5.36</td>
</tr>
</tbody>
</table>

**cause**

British National Corpus freq = 20207
Hatzivassiloglou & McKeown 1997

1. Build training set: label all adj. with frequency > 20; test agreement with human annotators
2. Extract all conjoined adjectives

Examples of conjoined adjectives:
- nice and comfortable
- nice and scenic
3. A supervised learning algorithm builds a graph of adjectives linked by the same or different semantic orientation.
4. A clustering algorithm partitions the adjectives into two subsets.
Pang, Lee, Vaithyanathan

- Movie review classification using Naïve Bayes, Maximum Entropy, SVM
  - Results do not reach levels achieved in topic categorization
- Various feature combinations (unigram, bigram, POS, text position)
  - Unigram presence works best
- Challenge: discourse structure
- Observation: subjectivity comes in many (low-frequency) forms → better to have more data
- Boot-strapping produces cheap data
- High-precision classifiers look for sentences that can be labeled subjective/objective with confidence
- Extraction pattern learner gathers patterns biased towards subjective texts
- Learned patterns are fed back into high precision classifiers
Sentiment Classification

**Features Usually Considered**

**A. Syntatic Features**

- What is syntactic feature? - Usage of principles and rules for constructing sentences in natural languages [wikipedia]
- Different usage of syntactic features:

<table>
<thead>
<tr>
<th>POS+punctuation</th>
<th>POS Pattern</th>
<th>Modifiers</th>
</tr>
</thead>
</table>

- **POS (part-of-speech) tags & punctuation**
  - eg: a sentence containing an adjective and “!” could indicate existence of an opinion “the book is great!”

- **POS n-gram patterns**
  - patterns like “n+aj” (noun followed by +ve adjective) - represents positive sentiment
  - patterns like “n+dj” (noun followed by -ve adjective) - express

- **Used a set of modifier features (e.g., very, mostly, not)**
  - the presence of these features indicate the presence of appraisal
  - “the book is not great”
  - “..this camera is very handy”
## Sentiment Classification

### Features Usually Considered

**B. Semantic Features**

- Leverage meaning of words
- Can be done manually/semi/fully automatically

<table>
<thead>
<tr>
<th>Score Based - Turney [2002]</th>
<th>Lexicon Based - Whitelaw et al. [2005]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use PMI calculation to compute the SO score for each word/phrase</td>
<td>Use appraisal groups for assigning semantics to words/phrases</td>
</tr>
<tr>
<td><strong>Idea:</strong> If a phrase has better association with the word “Excellent” than with “Poor” it is positively oriented and vice versa</td>
<td>Each phrase/word is manually classified into various appraisal classes (eg: deny, endorse, affirm)</td>
</tr>
</tbody>
</table>

Score computation:

\[
\text{PMI (phrase} \wedge \text{“excellent”)} - \text{PMI (phrase} \wedge \text{“poor”)}
\]
Sentiment Classification

**Features Usually Considered**

C. Link-Based Features
- Use link/citation analysis to determine sentiments of documents
- Efron [2004] found that opinion Web pages heavily linking to each other often share similar sentiments
- Not a popular approach

D. Stylistic Features
- Incorporate stylometric/authorship studies into sentiment classification
- Style markers have been shown highly prevalent in Web discourse [Abbasi and Chen 2005; Zheng et al. 2006; Schler et al. 2006]
- Ex.

<table>
<thead>
<tr>
<th>Authorship Style of Students</th>
<th>Authorship Style of Professors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colloquial@spoken style of writing</td>
<td>Formal way of writing</td>
</tr>
<tr>
<td>Improper punctuation</td>
<td>Proper punctuation</td>
</tr>
<tr>
<td>More usage of curse words &amp; abbreviations</td>
<td>Limited curse words &amp; abbreviation</td>
</tr>
</tbody>
</table>
Sentiment Classification
Abbasi, Chen & Salem (TOIS-08)

• Propose:
  • sentiment analysis of web forum opinions in multiple languages (English and Arabic)

• Motivation:
  • Limited work on sentiment analysis on Web forums
  • Most studies have focused on sentiment classification of a single language
  • Almost no usage of stylistic feature categories
  • Little emphasis has been placed on feature reduction/selection techniques

• New:
  • Usage of stylistic and syntactic features of English and Arabic
  • Introduced new feature selection algorithm: entropy weighted genetic algorithm (EWGA)
    • EWGA outperforms no feature selection baseline, GA and Information Gain
  • Results, using SVM indicate a high level of classification accuracy
## Sentiment Classification

Abbasi, Chen & Salem (TOIS-08)

<table>
<thead>
<tr>
<th>Category</th>
<th>Feature Group</th>
<th>English</th>
<th>Arabic</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syntactic</td>
<td>POS N-grams</td>
<td>varies</td>
<td>—</td>
<td>frequency of part-of-speech tags (e.g., NP_VB)</td>
</tr>
<tr>
<td></td>
<td>Word Roots</td>
<td>—</td>
<td>varies</td>
<td>frequency of roots (e.g., slm, ktb)</td>
</tr>
<tr>
<td></td>
<td>Word N-grams</td>
<td>varies</td>
<td>varies</td>
<td>word n-grams (e.g. senior editor, editor in chief)</td>
</tr>
<tr>
<td></td>
<td>Punctuation</td>
<td>8</td>
<td>12</td>
<td>occurrence of punctuation marks (e.g., !;…?)</td>
</tr>
<tr>
<td>Stylistic</td>
<td>Letter N-Grams</td>
<td>26</td>
<td>36</td>
<td>frequency of letters (e.g., a, b, c)</td>
</tr>
<tr>
<td></td>
<td>Char. N-grams</td>
<td>varies</td>
<td>varies</td>
<td>character n-grams (e.g., abo, out, ut, ab)</td>
</tr>
<tr>
<td></td>
<td>Word Lexical</td>
<td>8</td>
<td>8</td>
<td>total words, % char. per word</td>
</tr>
<tr>
<td></td>
<td>Char. Lexical</td>
<td>8</td>
<td>§</td>
<td>total char., % char. per message</td>
</tr>
<tr>
<td></td>
<td>Word Length</td>
<td>20</td>
<td>20</td>
<td>frequency distribution of 1–20-letter words</td>
</tr>
<tr>
<td></td>
<td>Vocab. Richness</td>
<td>8</td>
<td>8</td>
<td>richness (e.g., hapax legomena, Yule’s K)</td>
</tr>
<tr>
<td></td>
<td>Special Char.</td>
<td>20</td>
<td>21</td>
<td>occurrence of special char. (e.g., @#$%^&amp;*+)</td>
</tr>
<tr>
<td></td>
<td>Digit N-Grams</td>
<td>varies</td>
<td>varies</td>
<td>frequency of digits (e.g., 100, 17, 5)</td>
</tr>
<tr>
<td></td>
<td>Structural</td>
<td>14</td>
<td>14</td>
<td>has greeting, has url, quoted content, etc.</td>
</tr>
<tr>
<td></td>
<td>Function Words</td>
<td>250</td>
<td>200</td>
<td>frequency of function words (e.g., of, for, to)</td>
</tr>
</tbody>
</table>
## Sentiment Classification

Abbasi, Chen & Salem (TOIS-08)

**Table VII. Experiment 1b Results**

<table>
<thead>
<tr>
<th>Techniques</th>
<th>10-Fold CV</th>
<th>Bootstrap</th>
<th>Std. Dev.</th>
<th># Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base</td>
<td>87.95%</td>
<td>88.05%</td>
<td>4.133</td>
<td>26,870</td>
</tr>
<tr>
<td>IG</td>
<td>92.50%</td>
<td>92.08%</td>
<td>2.523</td>
<td>2,316</td>
</tr>
<tr>
<td>GA</td>
<td>92.55%</td>
<td>92.29%</td>
<td>2.893</td>
<td>2,017</td>
</tr>
<tr>
<td>SVMW</td>
<td>92.86%</td>
<td>92.34%</td>
<td>2.080</td>
<td>2,000</td>
</tr>
<tr>
<td>EWGA</td>
<td><strong>95.55%</strong></td>
<td><strong>95.06%</strong></td>
<td>2.969</td>
<td><strong>1,752</strong></td>
</tr>
<tr>
<td>Whitelaw et al. 2005</td>
<td>90.20%</td>
<td>-</td>
<td>-</td>
<td>49,911</td>
</tr>
<tr>
<td>Pang &amp; Lee 2004</td>
<td>87.20%</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Mullen &amp; Collier 2004*</td>
<td>86.00%</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Pang et al., 2002*</td>
<td>82.90%</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
## Sentiment Classification

Abbasi, Chen & Salem (TOIS-08)

### Table X. Characteristics of English and Arabic Test Bed

<table>
<thead>
<tr>
<th>Features</th>
<th>U.S. Forum</th>
<th></th>
<th></th>
<th># Features</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10-Fold CV</td>
<td>Bootstrap</td>
<td>Standard Dev.</td>
<td></td>
</tr>
<tr>
<td>Stylistic</td>
<td>71.40%</td>
<td>71.07%</td>
<td>3.324</td>
<td>867</td>
</tr>
<tr>
<td>Syntactic</td>
<td>87.00%</td>
<td>87.13%</td>
<td>2.439</td>
<td>12,014</td>
</tr>
<tr>
<td><strong>Stylistic + Syntactic</strong></td>
<td><strong>90.63%</strong></td>
<td><strong>90.59%</strong></td>
<td><strong>2.042</strong></td>
<td><strong>12,881</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Features</th>
<th>Middle Eastern Forum</th>
<th></th>
<th></th>
<th># Features</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accuracy</td>
<td>Bootstrap</td>
<td>Standard Dev.</td>
<td></td>
</tr>
<tr>
<td>Stylistic</td>
<td>80.20%</td>
<td>80.01%</td>
<td>4.145</td>
<td>1,166</td>
</tr>
<tr>
<td>Syntactic</td>
<td>85.42%</td>
<td>85.23%</td>
<td>2.457</td>
<td>12,645</td>
</tr>
<tr>
<td><strong>Stylistic + Syntactic</strong></td>
<td><strong>90.81%</strong></td>
<td><strong>90.69%</strong></td>
<td><strong>2.093</strong></td>
<td><strong>13,811</strong></td>
</tr>
</tbody>
</table>
Accurate Sentiment analysis on informal genre is important
Approach I: Linguistic-based Approach

- Preprocessing
  - Normalize URLs, user names (@URL, @USER)
  - Negation words (negation words -> NOT)
  - Slang words (LOL -> laugh out loud)
  - Spelling correction using WordNet (coooool -> cool)

- Target and issue detection
  - Entity Recognition System
    - “Ron Paul”, “Barack Obama”, “Mitt Romney”, etc.,
  - Mine Issue related words from Wikipedia
    - 64 key phrases for “Economic” and 27 for “Foreign Policy”
Approach I: Linguistic-based Approach

- Sentiment classification
  - Features
    - N-grams: All unique unigrams, bigrams and trigrams
    - POS tags: Part-Of-Speech tags generated by Stanford Parser
    - Gazetteer: SentiWordNet (Baccianella et al., 2010), Subjectivity Lexicon (Wiebe et al., 2004), Inquirer (Stone et al., 1966), Taboada (Taboada and Grieve, 2004), UICLexiconn (Hu and Liu, 2004) and LIWCLexicon(Pennebaker et al., 2001)
    - Word Cluster: WordNet synset information to expand the entries of each gazetteer
  - Punctuation and Capitalization
  - Binary classification problem
    - Characteristic of debate
Motivation I

- Target dependent sentiment analysis is not enough
  - Some domains need to detect issues (topics)
  - For example, a user comments on “Barack Obama” on his “Economics” stance

--"I agree 100%, Ron Paul, 30 years constitutional voting record, been right about every financial disaster that no one else in DC seems to have a singular clue about.

--We need a "peace through strength" attitude like Reagan had when he defeated the Soviet Union in the Cold War. Ron Paul would have caused us to lose that war."
Motivation II

- Tackle challenges from cross-genre approach
  - Movie reviews, tweets and forums have huge differences
  - long-tailed distribution of lexicon coverage
Approach II: With Global Social Features

- Social Cognitive Hypothesis

  - Trend - Indicative target-issue pairs
    - The public have biased sentiments on some target-issue pair
    - For example: “Obama, Economics” -> Negative

  - Comparison - Indicative target-target pairs

  - Consistency - User-target-issue consistency

(Hamilton and Sherman, 1996)
**Approach II: With Global Social Features**

- **Social Cognitive Hypothesis**
  - **Trend - Indicative target-issuse pair**
  - **Comparison - Indicative target-target pair**
    - Public have biased sentiment on some target-target pairs
    - For example, talking about “Obama” and “McCain” together usually indicate positive sentiment about “Obama”.

(Mason and Marca, 2004)
Approach II: With Global Social Features

- Social Cognitive Hypothesis
  - Trend - Indicative target-issue pair
  - Comparison - Indicative target-target pair
  - Consistency - User-target-issue consistency
    - Most confident label propagation
    - Majority voting
    - Weighted majority voting

--"I agree 100%, Ron Paul, 30 years constitutional voting record, been right about every financial disaster that no one else in DC seems to have a singular clue about.

--We need a "peace through strength" attitude like Reagan had when he defeated the Soviet Union in the Cold War. Ron Paul would have caused us to lose that war.

(Heider, 1946); (Situngrk and Khanafiah, 2004)
**Experiments Setup**

- **Data Collection**
  - *The tweet data set was automatically collected by hashtags*
    - #Obama2012 or #GOP2012 for positive tweets
    - #Obamafail or #GOPfail for negative tweets
    - filtered all tweets where the hashtags of interest were not located at the very end of the message
  - *The discussion forum data set was adapted from the “Election & Campaigns” board of a political forum.*

- **Training/Test Data**
  - 11382 Movie Reviews, 4646 tweets and 762 forum posts.
  - Three-fold cross-validation
  - Balanced data for training and testing -> naïve baseline is 50%.

- **Evaluation Metric: Accuracy**
Experin

<table>
<thead>
<tr>
<th>Features</th>
<th>Forum</th>
<th>Tweets</th>
<th>Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unigram</td>
<td>54.3%</td>
<td>81.6%</td>
<td>75.0%</td>
</tr>
<tr>
<td>Bigram</td>
<td>58.9%</td>
<td>79.3%</td>
<td>70.6%</td>
</tr>
<tr>
<td>Unigram + Bigram</td>
<td>58.2%</td>
<td>83.7%</td>
<td>75.8%</td>
</tr>
<tr>
<td>Unigram + Trigram</td>
<td>58.3%</td>
<td>84.0%</td>
<td>75.6%</td>
</tr>
<tr>
<td>Bigram + Trigram</td>
<td>59.6%</td>
<td>79.7%</td>
<td>69.7%</td>
</tr>
</tbody>
</table>
## Experimental Results II

<table>
<thead>
<tr>
<th>Approach</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>59.61%</td>
</tr>
<tr>
<td>+ Hypothesis 1</td>
<td>62.89%</td>
</tr>
<tr>
<td>+ Hypothesis 2</td>
<td>62.64%</td>
</tr>
<tr>
<td>+ Hypothesis 3</td>
<td>67.24%</td>
</tr>
<tr>
<td>+ Hypothesis 1+2</td>
<td>64.21%</td>
</tr>
<tr>
<td>+ Hypothesis 1+2+3</td>
<td>71.97%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Approach</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>59.61%</td>
</tr>
<tr>
<td>Baseline + most confident label propagation</td>
<td>62.89%</td>
</tr>
<tr>
<td>Baseline + Majority voting</td>
<td>62.64%</td>
</tr>
<tr>
<td>Baseline + Weighted voting</td>
<td>67.24%</td>
</tr>
</tbody>
</table>
Experimental Results II
Remaining Challenges I

- **Sarcasm**
  - “LOL..remember Obama chastising business’s for going to Vegas. Vegas would have cost a third of what these locations costs. But hey, no big deal... ”

- **Domain-specific Latent Sentiments**
  - “tell me how the big government, big bank backing, war mongering Obama differs from Bush?”. 
Multiple Sentiments:
- “....As a huge Ron Paul fan I have my disagreements with him.......but even if you disagree with his foreign policy.......the guy is spot on with everything and anything else.....”

Thread Structure:
- Performing sentiment analysis at post level, without taking into account the thread context (agree and disagree in reply relationship) might lead to errors.
OUTLINE

• Emotion Detection
• Subjectivity Overview
• Sentiment Analysis
• Opinion Mining
Opinion Retrieval

- Is the task of retrieving documents according to topic and ranking them according to opinions about the topic.

- Important when you need people’s opinion on certain topic or need to make a decision, based on opinions from others.

<table>
<thead>
<tr>
<th>General Search</th>
<th>Opinion Search</th>
</tr>
</thead>
<tbody>
<tr>
<td>search for facts</td>
<td>search for opinions/opinionated topics</td>
</tr>
<tr>
<td>rank pages according to some authority and relevance scores</td>
<td>Rank is based on relevance to topic and content of opinion</td>
</tr>
</tbody>
</table>
Opinion Retrieval

• “Opinion retrieval” started with the work of Hurst and Nigam (2004)
  - **Key Idea**: Fuse together topicality and polarity judgment ➔
    “opinion retrieval”
  - **Motivation**: To enable IR systems to select content based on a
certain opinion about a certain topic
  - **Method**:
    - **Topicality judgment**: statistical machine learning classifier (Winnow)
    - **Polarity judgment**: shallow NLP techniques (lexicon based)
  - No notion of ranking strategy
## Opinion Retrieval

### Summary of TREC Blog Track (2006 – 2008)

<table>
<thead>
<tr>
<th>TREC 2006 Blog Track</th>
<th>TREC 2007 Blog Track</th>
<th>TREC 2008 Blog Track</th>
</tr>
</thead>
</table>
| Opinion Retrieval    | Same as 2006 with 2 new tasks:  
- blog distillation (feed search)  
- polarity determination | Same as 2006 and 2007 with 1 new task:  
- Baseline blog post retrieval task (i.e. “Find me blog posts about X.”) |
"Well, yes, we could read your blog.... or you could just tell us about your school day."
Opinion Retrieval

Summary of TREC-2006 Blog Track [6]

- TREC-2006 Blog Track – Focus is on Opinion Retrieval
  - 14 participants
  - Baseline System
    - A standard IR system without any opinion finding layer
  - Most participants use a 2 stage approach

![](image)

STAGE 1

- Query
- Standard retrieval & ranking scheme
  - tf*idf
  - language model
  - probabilistic
- Ranked documents

STAGE 2

- Ranked opinionated documents
- Opinion related re-ranking/filter
  - dictionary based
  - text classification
  - linguistics
Opinion Retrieval

Summary of TREC-2006 Blog Track [6]

• TREC-2006 Blog Track
  • The two stage approach
    First stage
    • documents are ranked based on topical relevance
    • mostly off-the-shelf retrieval systems and weighting models
      • TF*IDF ranking scheme
      • language modeling approaches
      • probabilistic approaches.
    Second stage
    • results re-ranked or filtered by applying one or more heuristics for detecting opinions
    • Most approaches use linear combination of relevance score and opinion score to rank documents. eg:

\[ \alpha \cdot Score_{rel} + \beta \cdot Score_{opn} \]
Opinion Retrieval

Summary of TREC-2006 Blog Track [6]

Opinion detection approaches used:

• Lexicon-based approach [2,3,4,5]:
  (a) Some used frequency of certain terms to rank documents
      ..of greatest quality ...........
      ...... nice ......
      wonderful ......
      ... good battery life ...
  (b) Some combined terms from (a) with information about the
      • distance between sentiment words &
      • occurrence of query words in the document

Ex:
Query: “Barack Obama”
Sentiment Terms: great, good, perfect, terrific…
Document: “Obama is a great leader….”

• success of lexicon based approach varied
Opinion Retrieval

*Summary of TREC-2006 Blog Track [6]*

**Opinion detection approaches used:**

- **Text classification approach:**
  - training data:
    - sources known to contain opinionated content (eg: product reviews)
    - sources assumed to contain little opinionated content (eg: news, encyclopedia)
  - classifier preference: Support Vector Machines
  - **Features (discussed in sentiment classification section):**
    - n-grams of words—eg: beautiful/<ww>, the/worst, love/it
    - part-of-speech tags
  - Success of this approach was limited
    - Due to differences between training data and actual opinionated content in blog posts

- **Shallow linguistic approach:**
  - frequency of pronouns (eg: I, you, she) or adjectives (eg: great, tall, nice) as indicators of opinionated content
  - success of this approach was also limited
Opinion Retrieval
Highlight: Gilad Mishne (TREC 2006)

Gilad Mishne (TREC 2006)

- **Propose**: multiple ranking strategies for opinion retrieval in blogs

- Introduced 3 aspects to opinion retrieval
  - **topical relevance**
    - degree to which the post deals with the given topic
  - **opinion expression**
    - given a “topically-relevant” blog post, to what degree it contains subjective (opinionated) information about it
  - **post quality**
    - estimate of the quality of a “blog post”
    - assumption - higher-quality posts are likely to contain meaningful opinions
Opinion Retrieval
Highlight: Gilad Mishne (TREC 2006)

Step 1: Is blog post Relevant to Topic?

- Query
  - Language modeling based retrieval
  - Blind relevance feedback [9]
  - Topic relevance improvements
  - Term proximity [10,11]
  - Temporal properties

- Ranked documents

- Add terms to original query by comparing language model of top-retrieved docs ➔ entire collection
  - Limited to 3 terms

<table>
<thead>
<tr>
<th>Topic</th>
<th>Added terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>859. letting india into the club?</td>
<td>nuclear, times, friedman</td>
</tr>
<tr>
<td>867. cheney hunting</td>
<td>vice, dick, accident</td>
</tr>
<tr>
<td>896. global warming</td>
<td>climate, change, greenhouse</td>
</tr>
</tbody>
</table>

- Every word n-gram from the query treated as a phrase

- Determine if query is looking for recent posts
  - Boost scores of posts published close to time of the query date
Opinion Retrieval
Highlight: Gilad Mishne (TREC 2006)

Step 2: Is blog post Opinionated?
• Lexicon-based method- using GeneralInquirer
• GeneralInquirer
  • large-scale, manually-constructed lexicon
  • assigns a wide range of categories to more than 10,000 English words

• Example of word categories :
  • emotional category: pleasure, pain, feel, arousal, regret
  • pronoun category: self, our, and you;

“The meaning of a word is its use in the language” - Ludwig Wittgenstein (1958)
Opinion Retrieval
Highlight: Gilad Mishne (TREC 2006)

Step 2: Is blog post Opinionated?
For each post calculate two sentiment related values known as “opinion level”

Blog post → Calculate opinion level → “post opinion” level

Extract “topical sentences” from post to count the opinion-bearing words in it

Topical sentences:
> sentences relevant to topic
> sentences immediately surrounding them

Idea:
Feeds containing a fair amount of opinions are more likely to express an opinion in any of its posts

Method:
> use entire feed to which the post belongs
> topic-independent score per feed estimates the degree to which it contains opinions (about any topic)
Opinion Retrieval
Highlight: Gilad Mishne (TREC 2006)

Step 3: Is the blog post of good quality?
- A. Authority of blog post: Link-based Authority
  - Estimates authority of documents using analysis of the link structure

- Key Idea:
  - placing a link to a page other than your own is like “recommending” that page
  - similar to document citation in the academic world

- Follow Upstill et al (ADCS 2003) - inbound link degree (indegree) as an approximation
  - captures how many links there are to a page

- Post’s authority estimation is based on:
  - Indegree of a post p & indegree of post p’s feed

\[
\text{Authority} = \log(\text{indegree}=3)
\]
Opinion Retrieval
Highlight: Gilad Mishne (TREC 2006)

Step 3: Is the blog post of good quality?
  • B. Spam Likelihood
    • **Method 1**: machine-learning approach - SVM
    • **Method 2**: text-level compressibility - Ntoulas et al (WWW 2006)
      • Determine: How likely is post P from feed F a SPAM entry?
      • Intuition: Many spam blogs use “keyword stuffing”
        • High concentration of certain words
        • Words are repeated hundreds of times in the same post and across feed
        • When you detect spam post and compress them $\Rightarrow$ high compression ratios for these feeds
        • Higher the compression ratio for feed F, more likely that post P is splog (Spam Blog)
          \[ \text{comp. ratio} = \frac{\text{size of uncompressed pg.}}{\text{size of compressed pg.}} \]

Final spam likelihood estimate:

\[ \text{(SVM prediction)} \times \text{(compressibility prediction)} \]
Opinion Retrieval
Highlight: Gilad Mishne (TREC 2006)

Step 4: Linear Model Combination

1. Topic relevance language model
   - Top 1000 posts
   - Term proximity query [10,11]
   - Blind relevance feedback [9]
   - Temporal properties

2. Opinion level
   - “Feed opinion” level
   - “Post opinion” level

3. Post Quality
   - Part “Post quality” scores
   - Spam likelihood [13]
   - Link-based authority [12]

4. Weighted linear combination of scores
   - Partial “Post quality” scores
   - Partial “Opinion level” scores

Final scores

Ranked opinionated posts
Opinion Retrieval

*Summary of TREC-2008 Blog Track*

- Opinion Retrieval & Sentiment Classification
  - Basic techniques are similar – classification vs. lexicon
  - More use of external information source
    - Traditional source: WordNet, SentiWordNet
    - New source: Wikipedia, Google search result, Amazon, Opinion web sites (Epinions.com, Rateitall.com)
Opinion Retrieval

*Summary of TREC-2008 Blog Track*

- Retrieve ‘good’ blog posts
  1. **Expert search techniques**
     - Limit search space by joining data by criteria
     - Used characteristics: number of comments, post length, the posting time
       -> estimate strength of association between a post and a blog.
  2. **Use of folksonomies**
     - Folksonomy: collaborative tagging, social indexing.
       - User generating taxonomy
       - Creating and managing tagging.
     - Showed limited performance improvement
Opinion Retrieval

Summary of TREC-2008 Blog Track

• Retrieve ‘good’ blog posts

3. Temporal evidence
   ▪ Some investigated use of temporal span and temporal dispersion.
   ▪ Recurring interest, new post.

4. Blog relevancy approach
   • Assumption:
     • A blog that has many relevant posts is more relevant.
     • The top N posts best represent the topic of the blog
   • Compute two scores to score a given blog.
     • The first score is the average score of all posts in the blog
     • The second score is the average score of the top N posts that have the highest relevance scores.
   • Topic relevance score of each post is calculated using a language modeling approach.
Opinion Retrieval – Recent Work

He et al (CIKM-08)

• Motivation:
  • Current methods require manual effort or external resources for opinion detection

• Propose:
  • Dictionary based statistical approach - automatically derive evidence of subjectivity
    • Automatic dictionary generation – remove too frequent or few terms with skewed query model
    • assign weights – how opinionated. divergence from randomness (DFR)
      • For term w, divergence \(D(\text{opREl})\) from \(D(\text{Rel})\)
        ( Retrieved Doc = Rel + nonRel. \(\text{Rel} = \text{opRel} + \text{nonOpRel}\) )
    • assign opinion score to each document using top weighted terms
    • Linear combine opinion score with initial relevance score

• Results:
  • Significant improvement over best TREC baseline
  • Computationally inexpensive compared to NLP techniques
Zhang and Ye (SIGIR 08)

- **Motivation:**
  - Current ranking uses only linear combination of scores
  - Lack theoretical foundation and careful analysis
  - Too specific (like restricted to domain of blogs)

- **Propose:**
  - Generation model that unifies **topic-relevance** and **opinion generation** by a quadratic combination
    - Relevance ranking serves as weighting factor to lexicon based sentiment ranking function
  
  - Different than the popular linear combination
Zhang and Ye (SIGIR 08)

- Key Idea:
  - Traditional document generation model:
    - Given a query $q$, how well the document $d$ “fits” the query $q$
    - estimate posterior probability $p(d|q)$
  - In this opinion retrieval model, new sentiment parameter, $S$ (latent variable) is introduced

$$p(d|q, s) = \frac{I_{op}(d, q, s)I_{rel}(d, q)}$$

$I_{op}(d, q, s)$: given query $q$, what is the probability that document $d$ generates a sentiment expression $s$

- Tested on TREC Blog datasets – observed significant improvement
Challenges in opinion mining

Summary of TREC Blog Track focus (2006 – 2008)

<table>
<thead>
<tr>
<th>TREC 2006 Blog Track</th>
<th>TREC 2007 Blog Track</th>
<th>TREC 2008 Blog Track</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opinion Retrieval</td>
<td>Same as 2006 with 2 new tasks: -blog distillation (feed search) -polarity determination</td>
<td>Same as 2006 and 2007 with 1 new task: -Baseline blog post retrieval task (i.e. “Find me blog posts about X.”)</td>
</tr>
</tbody>
</table>

Main Lessons Learnt from TREC 2006, 2007 & 2008:

*Good performance in opinion-finding is strongly dependent on finding as many relevant documents as possible regardless of their opinionated nature*
Many Opinion classification and retrieval system could not make improvements.
- Used same relevant document retrieval model. Evaluate the performance of opinion module.

<table>
<thead>
<tr>
<th>Group</th>
<th>ΔMAP of Mix</th>
<th>ΔMAP of Positive</th>
<th>ΔMAP of Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>KLE</td>
<td>4.86%</td>
<td>6.08%</td>
<td>3.51%</td>
</tr>
<tr>
<td>UoGTr</td>
<td>-3.77%</td>
<td>-4.62%</td>
<td>-2.76%</td>
</tr>
<tr>
<td>UWaterlooEng</td>
<td>-6.70%</td>
<td>-1.69%</td>
<td>-12.33%</td>
</tr>
<tr>
<td>UIC_IR_Group</td>
<td>-22.10%</td>
<td>2.12%</td>
<td>-49.60%</td>
</tr>
<tr>
<td>UTD_SLP_Lab</td>
<td>-22.96%</td>
<td>-17.51%</td>
<td>-29.23%</td>
</tr>
<tr>
<td>Fub</td>
<td>-55.26%</td>
<td>-59.81%</td>
<td>-50.18%</td>
</tr>
<tr>
<td>Tno</td>
<td>-76.42%</td>
<td>-75.93%</td>
<td>-77.02%</td>
</tr>
<tr>
<td>UniNE</td>
<td>-43.68%</td>
<td>-39.41%</td>
<td>-48.49%</td>
</tr>
</tbody>
</table>

\[ \Delta \text{MAP} = \text{Opinion finding MAP score with only retrieval system} - \text{Opinion finding MAP score with opinion module} \]

- A lot of margins to research!!!
Challenges in opinion mining

Highlight of TREC-2008 Blog Track

Lee et al.  Jia et al. (TREC 2008)

- **Propose:**
  - Method for query dependent opinion retrieval and sentiment classification

- **Motivation:**
  - Sentiments are expressed differently in different query. Similar to the Blitzer’s idea.

- **New:**
  - Use external web source to obtain positive and negative opinionated lexicons.

- **Key Idea:**
  - Objective words: Wikipedia, product specification part of Amazon.com
  - Subjective words: Reviews from Amazon.com, Rateitall.com and Epinions.com
    - Reviews rated 4 or 5 out of 5: positive words
    - Reviews rated 1 or 2 out of 5: negative words
  - Top ranked in Text Retrieval Conference.
Challenges in opinion mining

• Polarity terms are context sensitive.
  • Ex. Small can be good for ipod size, but can be bad for LCD monitor size.
  • Even in the same domain, use different words depending on target feature.
    • Ex. Long ‘ipod’ battery life vs. long ‘ipod’ loading time
  • Partially solved (query dependent sentiment classification)

• Implicit and complex opinion expressions
  • Rhetoric expression, metaphor, double negation.
  • Ex. The food was like a stone.
  • Need both good IR and NLP techniques for opinion mining.

• Cannot divide into pos/neg clearly
  • Not all opinions can be classified into two categories
  • Interpretation can be changed based on conditions.
  • Ex. 1) The battery life is ‘long’ if you do not use LCD a lot. (pos)
    2) The battery life is ‘short’ if you use LCD a lot. (neg)
    Current system classify the first one as positive and second one as negative.
    However, actually both are saying the same fact.
Opinion Retrieval – Summary

- Opinion Retrieval is a fairly broad area (*IR, sentiment classification, spam detection, opinion authority…etc*)
- Important:
  - Opinion search is different than general web search
- Opinion retrieval techniques
    - Some approaches in opinion ranking layer:
      - Sentiment: lexicon based, ML based, linguistics
      - Opinion authority - trustworthiness
      - Opinion spam likelihood
      - Expert search technique
      - Folksonomies
  - Opinion generation model
    - Unify topic-relevance and opinion generation into one model
- Future approach:
  - Use opinion as a feature or a dimension of more refined and complex search tasks
OUTLINE

- Subjectivity Overview
- Sentiment Analysis
- Opinion Mining
- Emotion Detection