Opinion mining
Introduction – facts and opinions

• Two main types of textual information on the Web.
  – Facts and Opinions

• Current search engines search for facts (assume they are true)
  – Facts can be expressed with topic keywords.

• Search engines do not search for opinions
  – Opinions are hard to express with a few keywords
    • How do people think of Motorola Cell phones?
  – Current search ranking strategy is not appropriate for opinion retrieval/search.
Introduction – user generated content

• Word-of-mouth on the Web
  – One can express personal experiences and opinions on almost anything, at review sites, forums, discussion groups, blogs ... (called the user generated content.)
  – They contain valuable information
  – Web/global scale: No longer – one’s circle of friends

• Our interest: to mine opinions expressed in the user-generated content
  – An intellectually very challenging problem.
  – Practically very useful.
Introduction – Applications

• **Businesses and organizations:** product and service benchmarking. Market intelligence.
  – Business spends a huge amount of money to find consumer sentiments and opinions.
  – Consultants, surveys and focused groups, etc

• **Individuals:** interested in other’s opinions when
  – Purchasing a product or using a service,
  – Finding opinions on political topics,

• **Ads placements:** Placing ads in the user-generated content
  – Place an ad when one praises a product.
  – Place an ad from a competitor if one criticizes a product.

• **Opinion retrieval/search:** providing general search for opinions
  – Predicting behaviours and trends in finance, medicine, politics
The challenge

• 81% of Internet users have done online research on a product 20% do so on a typical day
• among readers of online reviews between 73% and 87% report that reviews had a significant influence on their purchase
• consumers report being willing to pay from 20% to 99% more for a 5-star-rated item than a 4-star-rated item (the variance stems from what type of item or service is considered);
• 32% have provided a rating on a product, service, or person via an online ratings system, and 30% have posted an online comment or review regarding a product or service.
"Great location, hood value"

Reviewed February 26, 2012
3 people found this review helpful

Stayed for 2 nights. Able to walk from the train, but would also be an easy short taxi. Close to the centre of town and to public transport. Hotel provided us with travel cards that got us free transport on the local public system. Room is very clean and crisp, on the smaller side but it is Europe after all....
A truly remarkable camera. 08/12/2008

Advantages HD video recording, great photos, looks good
Disadvantages EXPENSIVE!

This camera is one hell of a camera and I was not disappointed buying this. I have been making some extra cash doing photography for parties and similar types of things just lately and this has been my new best friend. It is reliable and takes very high quality pictures as well as having the ability to record HD, yes! HDI, video at resolution 1080p (which is a lot better than most cameras). First I'd like to comment on the design, looking both professional as well as striking; this really gives off the impression that you are a professional photographer and it is a pleasure and a pride to take .... more

The mighty 5D (mark II) 06/12/2010

Advantages Image quality, video function, Canon lenses
Disadvantages overheating

This incredible camera I have owned now for about half a year. Its a reasonable weight to begin with. Well under 3kg with mounted lens. There are a great selection of Canon lenses to choose from too. The 35mm full frame sensor allows for zero cropping when using Canon lenses and along with 21MP creates very good quality photographs. With this camera you can program your own manual user settings and colour/saturation schemes while also altering the camera's top ISO speed as well for low light situations. The functions button are quite easy to navigate around and easier if you have previously ... more
WHY IS THIS USEFUL ON THE WEB?
Correlation between DJ and web mood ("calm" (GOMPS))
Correlation between events (political, social) and web mood
Frequency of negative terms in News headlines from July 31st to August 9, 2011
Main resources

• **Lexicons**
  - General Inquirer (Stone et al., 1966)
  - OpinionFinder lexicon (Wiebe & Riloff, 2005)
  - SentiWordNet (Esuli & Sebastiani, 2006)

• **Annotated corpora**
  - Used in statistical approaches (Hu & Liu 2004, Pang & Lee 2004)
  - MPQA corpus (Wiebe et. al, 2005)

• **Tools**
  - Algorithm based on minimum cuts (Pang & Lee, 2004)
  - OpinionFinder (Wiebe et. al, 2005)
sentiwordnet.isti.cnr.it/
SentiWordNet

Positive | Subjective | Negative

PN polarity

SO polarity

Objective

Term Sense Position
WordNet Affect Taxonomy

dislike
  |—— disgusting
  |     |—— repugnance
  |     |—— nausea
  |     |—— antipathy
  |     |—— contempt
  |—— disapproval
  |     |—— disinclination
  |     |—— alienation
  |     |     |—— isolation
  |     |—— unfriendliness
Loughran and McDonald Financial Sentiment Dictionaries

Updated: 2011

Note: We thank Cam Harvey and others who suggested some of the modifications we’ve included in these lists. The original versions of the word lists are described in Loughran and McDonald (Journal of Finance, V66, pp. 35-65, 2011). The word list files contain are comma delimited and contain a word followed by the version year. Not for commercial use without authorization. Copyright 2009.

- Negative Words
- Positive Words
- Uncertainty Words
- Litigious Words
- Modal Words Strong
- Modal Words Weak
- Download zip folder with all lists
- Download zip folder in WordStat format (contains .cat and .NFO files)
Other sentiment lexicons

- Bing Liu's Opinion Lexicon
- MPQA Subjectivity Lexicon
- SentiWordNet
- Harvard General Inquirer
- LIWC

- Downloadable from: http://sentiment.christopherpotts.net/lexicons.html#opinionlexicon
### Disagreements between polarity lexicons

Christopher Potts, [Sentiment Tutorial](#), 2011

<table>
<thead>
<tr>
<th></th>
<th>Opinion Lexicon</th>
<th>General Inquirer</th>
<th>SentiWordNet</th>
<th>LIWC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MPQA</strong></td>
<td>33/5402 (0.6%)</td>
<td>49/2867 (2%)</td>
<td>1127/4214 (27%)</td>
<td>12/363 (3%)</td>
</tr>
<tr>
<td><strong>Opinion Lexicon</strong></td>
<td>32/2411 (1%)</td>
<td>1004/3994 (25%)</td>
<td></td>
<td>9/403 (2%)</td>
</tr>
<tr>
<td><strong>General Inquirer</strong></td>
<td></td>
<td>520/2306 (23%)</td>
<td>1/204 (0.5%)</td>
<td></td>
</tr>
<tr>
<td><strong>SentiWordNet</strong></td>
<td></td>
<td></td>
<td>174/694 (25%)</td>
<td></td>
</tr>
<tr>
<td><strong>LIWC</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Annotated corpora

- [http://www.cyberemotions.eu/data.html](http://www.cyberemotions.eu/data.html)
- [http://www.di.unito.it/~tutreeb/sentiTUT.html](http://www.di.unito.it/~tutreeb/sentiTUT.html) (in italian)
- Stanford Twitter Corpus: [http://help.sentiment140.com/for-students](http://help.sentiment140.com/for-students)
- HCR and OMD datasets: [https://bitbucket.org/speriosu/updown](https://bitbucket.org/speriosu/updown)
- Sentiment Strength Corpora: [http://sentistrength.wlv.ac.uk/](http://sentistrength.wlv.ac.uk/)
- SemEval: [http://www.cs.york.ac.uk/semeval-2013/task2/](http://www.cs.york.ac.uk/semeval-2013/task2/)
Tools
Pulse of the Nation:
U.S. Mood Throughout the Day inferred from Twitter

http://www.ccs.neu.edu/home/amislove/twittermood
Twitter investor sentiment

Investor Sentiment
@Tweet_Sentiment

Helping investor navigate the Twitter Sentiment. AKA, using twitter to predict the predict a Bull or Bear market. Invest real-time using this twitter feed!

chocolatichipapps.com

Iscriviti a febbraio 2011

Tweet Tweet e risposte

Ritwittato da Investor Sentiment
Scott Eddy @MrScottEddy · 4 feb 2011
Congress Grills Facebook On Plans To Share User Addresses, Cell Numbers http://huff.to/eFC0Cu

Tweet @MrScottEddy · 4 feb 2011

Yahoo Finance @YahooFinance · 4 feb 2011
Stocks up ahead of US jobs data, Egypt woes loom http://y.hoo.it/hvDrCB #Futures

Tweet @YahooFinance · 4 feb 2011

Ritwittato da Investor Sentiment
QualityStocks @QualityStocks · 4 feb 2011
January jobs report forecast to show modest gains http://ow.ly/3Q0kN ~ http://disclaim.it/1f8ewa

Tweet @QualityStocks · 4 feb 2011

Investor Sentiment @Tweet_Sentiment · 4 feb 2011
dvolatile: RT @BreakingNews: Suspect in custody after gunman briefly hijacks Greyhound bus in N.C.; no one hurt - ... http://ff.im/xlcwh

Tweet @BreakingNews · 4 feb 2011
Discover what people are really saying on Twitter. With Twitratr you can distinguish negative from positive tweets surrounding a brand, product, person or topic.

<table>
<thead>
<tr>
<th>TERM</th>
<th>POSITIVE TWEETS</th>
<th>NEUTRAL TWEETS</th>
<th>NEGATIVE TWEETS</th>
<th>TOTAL TWEETS</th>
</tr>
</thead>
<tbody>
<tr>
<td>st ives</td>
<td>70</td>
<td>384</td>
<td>11</td>
<td>465</td>
</tr>
</tbody>
</table>

15.05% POSITIVE

1. i really want to love st ives apricot scrub, but it irritates my skin soo much. (view)
2. rt @kesiahosking: sunshine was smiling at you annie =)) rt @anniegreenwood st ives harbour basking in november sunshine http://flic.kr/p/8tk2sq (view)
3. sunshine was smiling at you annie =)) rt @anniegreenwood st ives harbour basking in november sunshine http://flic.kr/p/8tk2sq (view)
4. looking at st ives (uk:siv). great stats, but printing? (view)

82.58% NEUTRAL

1. @oldergirlbeauty GURL, I was all about the Aqua Net & the St. Ives liquid hairspray in the purple bottle. Where’s my banana clip? (view)
2. RT @inscriptions: Loved the final episode of Junior Masterchef! Alex will be at St Ives Village Sat 11th to show us a thing or two! (view)
3. Loved the final episode of Junior Masterchef! Alex from top12 is coming to St Ives Village Sat 11th to show us a thing or two about cooking! (view)
4. A Town On Canvas Called St Ives http://ning.fm/onNWi (view)

2.37% NEGATIVE

1. st. ives apricot scrub is bad for your face. you may not notice it but it scratches up your face and its bad... http://bit.ly/dttmci (view)
2. st ives face scrub receive negative comments. lots of it o.o (view)
3. @fandomonymous not sure how bad your acne is, but st. ives green tea cleanser works well on my skin. really cleans out my pores. (view)
4. sco prem: goal st ives city 2 towerhill blues 0 lucas k (43) (view)
5. sco prem: goal st ives city 1
RT @BuildYourLoveUp: RT @BuildYourLoveUp: I wish @itsimreeeee went to the same school as me. I miss my best friend, and almost everyone at Bonanza sucks. ;/

shout out to the helicopter circling our school this morning with a spot light.. I love Bonanza http://t.co/j2EDX0cS

RT @ADReamGONe: RT @ADReamGONe: Man, I love Bonanza Imfao.

RT @ADReamGONe: RT @ADReamGONe: Man, I love Bonanza Imfao.
Sentiment analysis for microsoft

Sentiment by Percent
- Positive (57%)
- Negative (43%)

Sentiment by Count
- Positive (59)
- Negative (45)

Tweets about: microsoft

lsaydumb: @Yopurn, in my humble opinion you have nothing to do on the @Xbox Live. What the fuck is @Microsoft doing?!
Posted 46 seconds ago

Megan Maracle: I hate this class. #Microsoft #die
Posted 2 minutes ago

dilwortha: @carasmith10 oh okay, you'll have to explain when i see you as i don't understand this disk haha. is it for microsoft project do you?
Posted 5 minutes ago

jlebrch: @rssldnphy it happens to be microsoft this time, but a superset is the next best thing from a compiled bytecode, as valid JS is also
http://www.sentiment140.com/

The results for this query are: Accurate
Twitter Sentiment Visualization

Keywords: Beppe Grillo

Sentiment viz
Tweet Sentiment Visualization

Beppe Grillo (213)

Keywords: Beppe Grillo
Opinion Finder

OpinionFinder 1.x Release Page

OpinionFinder 1.x Available versions

OpinionFinder 1.x relies on many external software packages (e.g. SUNDANCE, SCOL, BoosTexter) which are neither built nor supported by our group. Since OpinionFinder was originally released in 2005, there are some compatibility issues with versions of various software and packages. We have reports that these problems sometimes result in an exhausting and even unsuccessful installation process of OpinionFinder. Since many of the people involved in the original development have graduated and left the group, we do not currently have the resources to address these compatibility issues concerning the required external software packages. Although we do not have the resources to bring OpinionFinder 1.x fully up-to-date, we are currently working on a new version of OpinionFinder. OpinionFinder 2 is being written in Java and will be platform-independent.

LICENSE AGREEMENT

Version 1.5

- README - OpinionFinder 1.5
- Download OpinionFinder 1.5
Plenty of resources, but what about methods?

• Why is opinion search different from web search?
  – The Task
  – Sentiment Ranking
  – Visualization of results

• What are the methods?
Tasks: Three types of opinion search

1. **Direct Opinions**: sentiment expressions on some objects, e.g., products, events, topics, persons.
   - E.g., “the picture quality of this camera is great”
   - Subjective

2. **Comparisons**: relations expressing similarities or differences of more than one object. Usually expressing an ordering.
   - E.g., “car x is cheaper than car y.”
   - Objective or subjective.

3. **Predictions**: global “mood” of “some fragment” of the web is useful to predict relevant events in finance, medicine, commerce, politics
   - E.g. “calm”, “anxious”, “happy”
   - Subjective, but aggregated
1. Direct opinions:

- Find the opinion of a person or organization (opinion holder) on a particular object or a feature of the object.
  - E.g., what is Bill Clinton’s opinion on abortion?
- Find positive and/or negative opinions on a particular object (or some features of the object), e.g.,
  - customer opinions on a digital camera.
  - public opinions on a political topic.
1.1 Find the opinion of a person on X

- In some cases, the general search engine can handle it, i.e., using suitable keywords.
  - Obama’s opinion on free weapons

- Reason:
  - One person or organization usually has only one opinion on a particular topic.
  - The opinion is likely contained in a single document.
  - Thus, a good keyword query may be sufficient.
President-elect Obama's View on Nuclear Weapons - 33 Minutes
www.33-minutes.com/.../obama-nuclear-weapons....
Where does President-elect Obama stand on nuclear weapons and missile defense?

What is Barack Obama's view on nuclear weapons - Wiki Answers
wiki.answers.com › ... › Nuclear Weapons
Barack Obama does not support nuclear weapons. He wants elimination of all nuclear weapons in the world, saying that the United States should reduce its ...

Where Obama and Romney stand on gun control - The Washington ...
www.washingtonpost.com/.../obama..../gJQAwMpN...
20/lug/2012 – “I'm not going to take away your guns,” Obama promised in September the assault-weapons ban, reaffirming that position in a 2008 debate.

Obama calls for strengthening gun rules, Reid says no time for ...
www.foxnews.com/.../obama-talks-limiting-some-...
26/lug/2012 – As President Obama added his voice to the push for stricter gun rules in the wake of the ... He noted that Obama supports the reinstatement of the assault weapons ban which expired in 2004, but said he .... View All Polls ...

Assault Weapons Ban: Is Obama Softening His Position On Guns?
www.policymic.com/.../assault-weapons-b...
The president defended hunting and showed supporters of the Second Amendment some love during an interview with 'The New Republic.' Is he softening his ...
1.2 Find opinions on an object

We use product reviews as an example:

• Searching for opinions in product reviews is different from general Web search.
  – E.g., search for opinions on “Motorola RAZR V3”

• General Web search (for a fact): rank pages according to some authority and relevance scores.
  – The user views the first page (if the search is perfect).
  – One fact = Multiple facts

• Opinion search: rank is desirable, however
  – reading only the review ranked at the top is not appropriate because it is only the opinion of one person.
  – One opinion ≠ Multiple opinions
2. Comparisons:

- **Gradable**
  - **Non-Equal Gradable**: Relations of the type greater or less than
  - Ex: “optics of camera A is better than that of camera B”
  - **Equative**: Relations of the type equal to
  - Ex: “camera A and camera B both come in 7MP”
  - **Superlative**: Relations of the type greater or less than all others
  - Ex: “camera A is the cheapest camera available in market”
3. Predictions
Other issues on opinion search: **Ranking and result Visualization**

- **Opinion ranking:**
  - should produce two rankings
    - Positive opinions and negative opinions
    - Some kind of summary of both, e.g., # of each
  - Or, one ranking but
    - The top (say 30) reviews should reflect the natural distribution of all reviews (assume that there is no spam), i.e., with the right balance of positive and negative reviews.

- **Questions:**
  - Should the user read all the top reviews? OR
  - Should the system prepare a **summary** of the reviews?
Other issues on opinion search: **Ranking and result Visualization** (cont’d)

- Reviews can be regarded as traditional surveys.
  - In traditional survey, returned survey forms are treated as raw data.
  - Analysis is performed to summarize the survey results.
    - E.g., % against or for a particular issue, etc.

- In opinion search,
  - **Can a summary be produced?**
  - **What should the summary be?**
Example of result visualization

Twitter Sentiment

Sentiment analysis for Obama

Sentiment by Percent

Positive (62%)
Negative (38%)

Sentiment by Count

Positive (40)
Negative (25)
Another example of visualization

Ristoranti: Roma

<table>
<thead>
<tr>
<th>Cucina</th>
<th>Quartieri</th>
<th>Opzioni</th>
<th>Prez</th>
</tr>
</thead>
<tbody>
<tr>
<td>✔ Tutti (3.771)</td>
<td>✔ Tutti (3.771)</td>
<td>✔ Tutti (3.771)</td>
<td>✔ Tutti (3.771)</td>
</tr>
<tr>
<td>☐ Africana (13)</td>
<td>☐ Esquilino / Termini (38)</td>
<td>☐ Accetta prenotazioni (1.577)</td>
<td>☐</td>
</tr>
<tr>
<td>☐ Americana (23)</td>
<td>☐ Ripa / Aventino (4)</td>
<td>☐ Adatto ai bambini (292)</td>
<td>☐</td>
</tr>
<tr>
<td>☐ Asiatica (19)</td>
<td>☐ Sant’Angelo / Ghetto (18)</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>☐ Bar (33)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Recensioni attendibili

1-20 su 3.771

In ordine di [ Popolarità ] [ Prezzo ] ☐ Mostra prima i ristoranti prenotabili (23)

1 Il Cavalluccio Marino

N. 1 di 3.389 ristoranti in Roma

62 recensioni

Possibilità per mangiare: Accetta prenotazioni, Consegna, Dopo mezzanotte

" Da provare! " 17 apr 2012
" Sicuramente da provare " 16 apr 2012

2 Ad Hoc

N. 2 di 3.389 ristoranti in Roma

642 recensioni
Twitter Sentiment Visualization
Methods

• Why is opinion search different from web search?
  – Task
  – Ranking
  – Visualization

• What are the methods?
A formalization of the task

• Basic components of an opinion:
  • Opinion holder: The person or organization that holds a specific opinion on a particular object.
  • Object: on which an opinion is expressed (it can be described by features, e.g. for an hotel room: dimension, clean, silent, cost,..)
  • Opinion: a view, attitude, or appraisal on an object (or object feature) from an opinion holder.
Opinion mining “grain”

- At the document (or review) level:
  - Task: sentiment classification of reviews
  - Classes: positive, negative, and neutral
  - Assumption: each document (or review) focuses on a single object (not true in many discussion posts) and contains opinion from a single opinion holder.
  - Example: Movie reviews

- At the sentence level:
  - Task 1: identifying subjective/opinionated sentences
    - Classes: objective and subjective (opinionated)
  - Task 2: sentiment classification of sentences
    - Classes: positive, negative and neutral.
    - Assumption: a sentence contains only one opinion; not true in many cases.
    - Then we can also consider clauses or phrases.
  - Example: hotel reviews
Opinion Mining Tasks (cont.)

• At the feature level (Example: product reviews, usually you want know opinions on various features of the product to improve or to compare)
  • Task 1: Identify and extract object features that have been commented on by an opinion holder (e.g., a reviewer).
  • Task 2: Determine whether the opinions on the features are positive, negative or neutral.
  • Task 3: Group feature synonyms.

• Opinion holders: identify holders is also useful, e.g., in news articles, etc, but they are usually known in the user generated content, i.e., authors of the posts.
Approaches

• Machine learning
  – Naïve Bayes
  – Maximum Entropy Classifier
  – SVM
  – Markov Blanket Classifier (aka of Markov models)
    • Accounts for conditional feature dependencies
    • Allowed reduction of discriminating features from thousands of words to about 20

• Unsupervised methods
  • Use opinion lexicons

Assume pairwise independent features
Example of machine-learning classifier

- LingPipe (downloadable from http://alias-i.com/lingpipe/demos/tutorial/sentiment/read-me.html)
- Bo Pang and Lillian Lee: “A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts”
- Application: sentiment classification of movie reviews
LingPipe Polarity Classifier (1)

- First eliminate objective sentences, then use remaining sentences to classify document polarity (reduce noise)

- Two steps: i) subjective/objective classification; ii) positive/negative classification of subjective sentences
LingPipe Polarity Classifier (2)

• Basic idea is that each sentence should be classified as either subjective or objective, but the decision depends on two factors:
  – The result of an individual classification of the sentence, based on the words it contains (e.g. “I” “me” “think”..);
  – The classification of adjacent sentences: it is assumed that adjacent sentences should have same classification
LingPipe Polarity Classifier (3)

• Uses **unigram features (single words)** extracted from movie review annotated dataset and a **Naïve Bayes classifier** to assign each sentence $s_i$ a score $ind1(s_i)$ (probability $s_i$ is subjective, C1) and a score $ind2(s_i)$ (probability $s_i$ is objective, C2)

• Assumes that adjacent sentences are likely to have similar subjective-objective (SO) polarity.

• Given two sentence pairs $(s_i, s_j)$, $assoc(s_i, s_j)$ is a measure of their association that decreases with distance
LingPipe Polarity Classifier (4)

- Let $s_1..s_n$ be a set of sentences; associate a node to each sentence.
- Let $s$ and $t$ be additional nodes, $s=$source, $t=$sink
- Add $n$ edges from $s$ to each $s_i$ with weight $\text{ind}_1(s_i)$
- Add $n$ edges from each $s_i$ to $t$ with weight $\text{ind}_2(s_i)$
- Add $\binom{n}{2}$ edges $(s_i, s_j)$ with weight $\text{assoc}(s_i, s_j)$
LingPipe Polarity Classifier (5)

- Uses **a min-cut algorithm** to efficiently extract subjective sentences. Assign sj to C1 or C2 so as to minimize the partition cost of:

\[
\sum_{s_i \in C1} ind_1(s_i) + \sum_{s_i \in C2} ind_2(s_i) + \sum_{s_i \in C1, s_j \in C2} assoc(s_i, s_j)
\]
LingPipe Polarity Classifier (example)

Graph for classifying three items.

Example with 3 nodes, Y, M and N + source and sink. Edge weights computed as in step 1 of algorithm; min-cut is shown with a dashed line. Min-cut puts Y and M in C1, N in C2.
LingPipe Polarity Classifier (evaluation)

- Accurate as baseline but uses only 22% of content in test data (average)
- Metrics suggests properties of movie review structure

Figure 4: Accuracies using N-sentence extracts for NB (left) and SVM (right) default polarity classifiers.
Sentiment classification based on sentiment lexicons

• Untrained methods
• Use a priori, semantic knowledge on words expressing positive or negative attitudes
• Either general purpose or domain-dependent lexicon
• Combine weight of positive/negative words to assign polarity
SentiWordNet

• Based on WordNet “synsets”
  – http://wordnet.princeton.edu/

• Ternary classifier
  – Positive, negative, and neutral scores for each synset

• Provides means of gauging sentiment for a text
SentiWordNet: Construction

• Created training sets of synsets, \( L_p \) and \( L_n \)
  – Start with small number of synsets with fundamentally positive or negative semantics, e.g., “nice” and “nasty”
  – Use WordNet relations, e.g., direct antonymy, similarity, derived-from, to expand \( L_p \) and \( L_n \) over K iterations
  – \( L_o \) (objective) is set of synsets not in \( L_p \) or \( L_n \)

• Trained classifiers on training set
  – Rocchio and SVM
  – Use four values of K to create eight classifiers with different precision/recall characteristics (ensamble)
  – As K increases, P decreases and R increases
SentiWordNet: Results

• 24.6% synsets with Objective<1.0
  – Many terms are classified with some degree of subjectivity
• 10.45% with Objective<=0.5
• 0.56% with Objective<=0.125
  – Only a few terms are classified as definitively subjective
• Difficult (if not impossible) to accurately assess performance
• However as time passes, it gets more accurate
• Recently connected with Babelnet to obtain polarity lexica in many languages
SentiWordNet: How to use it

• Use score to select features (+/-)
  – e.g. Zhang and Zhang (2006) used words in corpus with subjectivity score of 0.5 or greater

• Combine pos/neg/objective scores to calculate document-level score
  – e.g. a simple sum

\[ p(s_i) = \sum_{w_i \in s_i, \text{SentiWordNet}} \text{weight}(w_i) \]
Opinion Mining research Roadmap