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!!
- Organization internal data
 - Customer feedback from emails, call centers, etc.
- News and reports
 - Opinions in news articles and commentaries

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.

The challenge

- People express opinions in complex ways
- In opinionated texts, lexical content alone can be misleading:
 - Intra-textual and sub-sentential reversals, negation, topic change common
 - Rhetorical devices/modes such as
 - sarcasm, irony, implication, etc.
- Honda Accords and Toyota Camrys are nice sedans
- Honda Accords and Toyota Camrys are nice sedans,
 but hardly the best car on the road

Dozen of companies and systems and thousands of research papers

















Applications in Business Intelligence: products reviews

- Question: "Why aren't consumers buying our laptop?"
- We know the concrete data: price, specs, competition, etc.
- We want to know subjective data: "the design is tacky," "customer service was condescending"
- Misperceptions are also important, e.g. "updated drivers aren't available" (even though they are)

Cross-domain applications

- Insights and applications from opinion mining have been useful in other areas
 - Politics/political science
 - Analyzing trends, identifying ideological bias, targeting advertising/messages, gauging reactions, etc.
 - Evaluation of public/voters' opinions
 - Other applications:
 - Law/policy making
 - Sociology (Modeling trust and influence in the blogosphere)
 - Psychology (dream sentiment analysis)

Tripadvisor

76 reviews from our community

Write a Review



Trip type

Family reviews (8)

Couples reviews (39)

Business reviews (6)

Solo travel reviews (6)

Friends reviews (5)

See which rooms travelers prefer - 3 traveler tips

76 reviews sorted by Date ▼

Rating

English first \$



"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....

Reviewer

Hobart, Australia

4 reviews

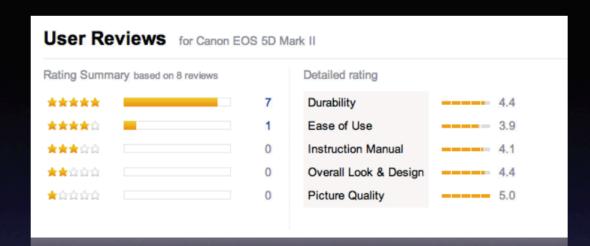
5 helpful votes

More ▼

Was this review helpful?

Problem with this review?

ciao.co.uk



★★★★ 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! HD!, 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

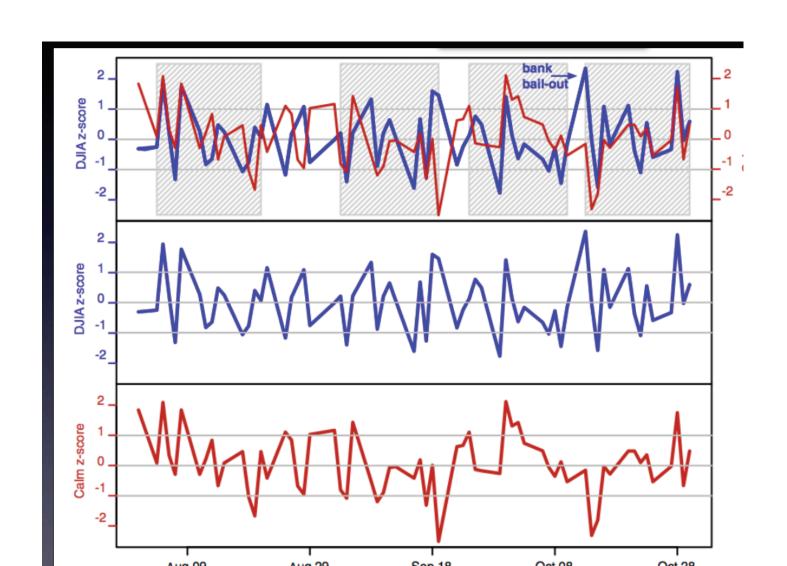
★★★★ The mighty 5D (mark II) 06/12/2010



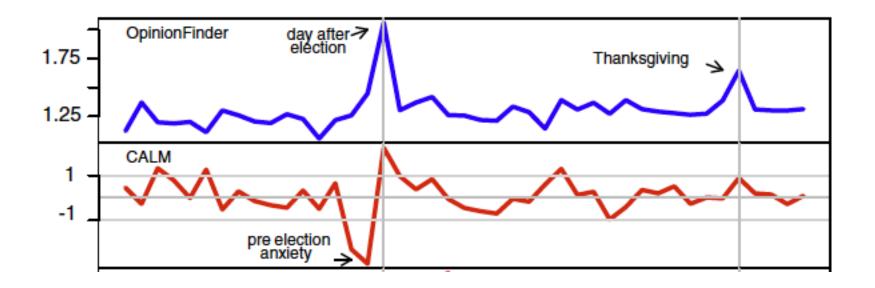
- Advantages Image quality, video function, Canon lenses
- Disadvantages overheating

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

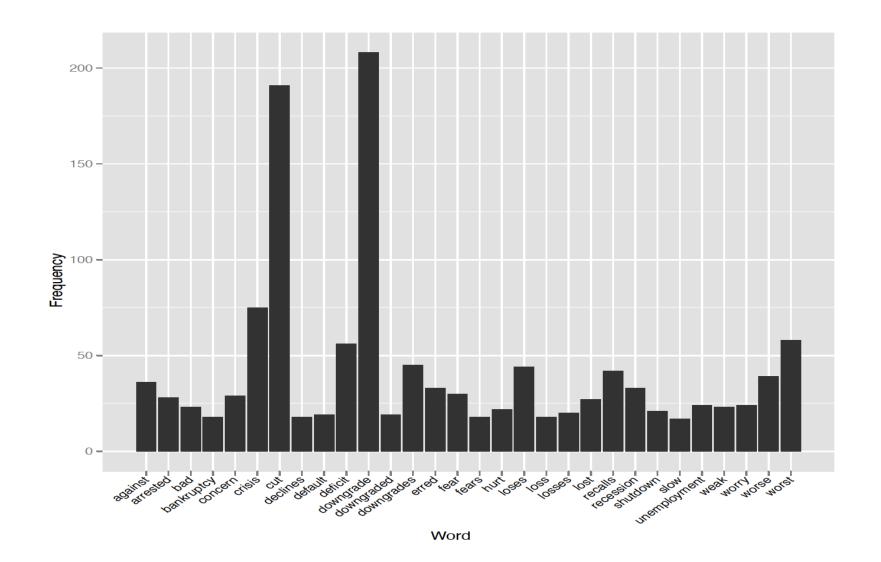
Correlation between DowJones and «web mood» (GOMPS, «calm» score)



Correlation between events (political, social) and web mood (Opinion Finder)



Frequency of negative terms in News headlines from July 31st to August 9, 2011



Main resources for Opinion Mining



Lexicons

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



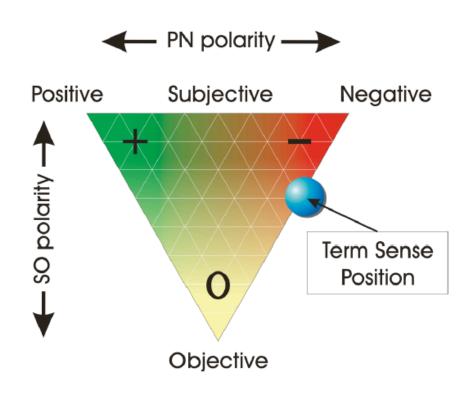
- Used in statistical approaches (Hu & Liu 2004, Pang & Lee 2004)
- MPQA corpus (Wiebe et. al, 2005)
- https://www.w3.org/community/sentiment/wiki/Datase ts
- Methods
 - Algorithm based on minimum cuts (Pang & Lee, 2004)
 - OpinionFinder (Wiebe et. al, 2005)
 - More recent deep-based algorithms
 (https://onlinelibrary.wiley.com/doi/abs/10.1002/widm.1253)



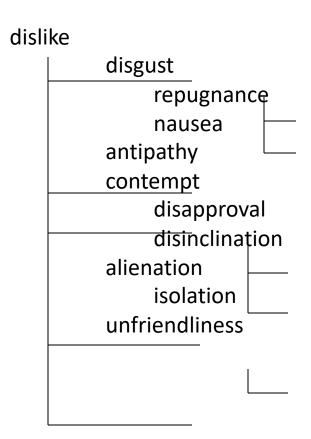


Lexicons





WordNet Affect Taxonomy



http://www.nd.edu/~mcdonald/Word _Lists.html

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 <u>Loughran and McDonald</u> (*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
- Harvard General Inquirer
- LIWC
- Downloadable from: http://sentiment.christopherpotts.net/lexicon s.html#opinionlexicon

Disagreements between polarity lexicons

Christopher Potts, Sentiment Tutorial, 2011

	Opinion Lexicon	General Inquirer	SentiWordNet	LIWC
MPQA	33/5402 (0.6%)	49/2867 (2 %)	1127/4214 (27%)	12/363 (3%)
Opinion Lexicon		32/2411 (1%)	1004/3994 (25%)	9/403 (2 %)
General Inquirer			520/2306 (23%)	1/204 (0.5%)
SentiWordNet				174/694 (25%)
LIWC				

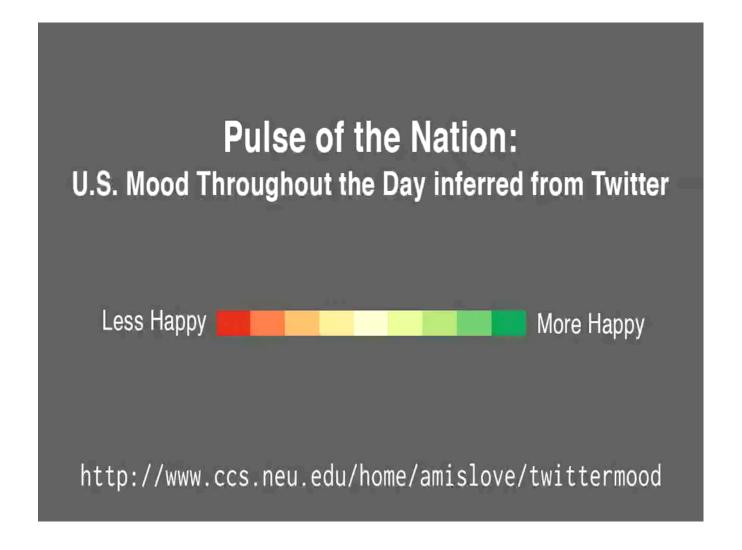
Annotated corpora



- http://www.cyberemotions.eu/data.html
- http://www.di.unito.it/~tutreeb/sentiTUT.html (in italian)
- Stanford Twitter Corpus: http://help.sentiment140.com/for-students
- HCR and OMD datasets: https://bitbucket.org/speriosu/updown
- Sentiment Strength Corpora: http://sentistrength.wlv.ac.uk/
- Sanders: http://www.sananalytics.com/lab/twitter-sentiment/
- SemEval: http://www.cs.york.ac.uk/semeval-2013/task2/

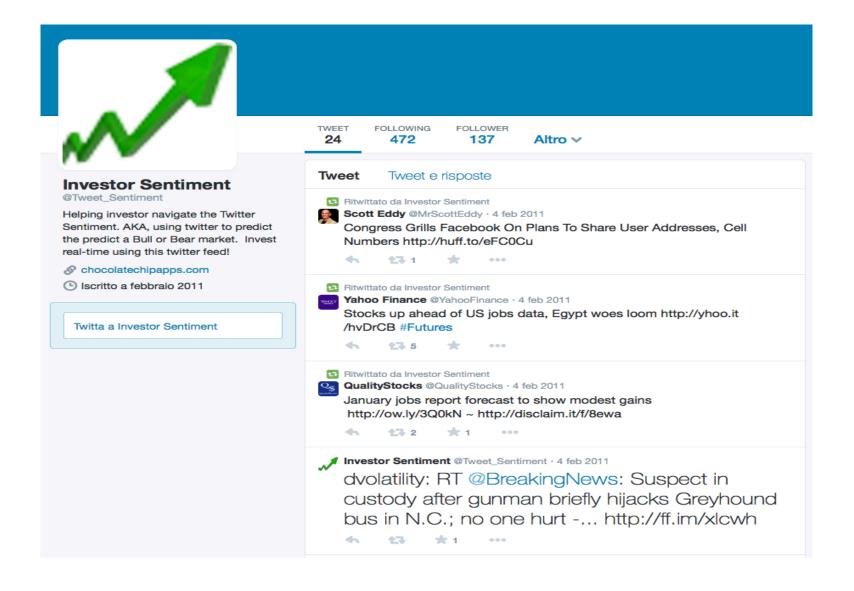
On-line tools





http://www.ccs.neu.edu/home/amislove/twittermood/

Twitter investor sentiment





twitrratr

SEARCH

Discover what people are really saying on Twitter. With Twitrratr you can distinguish negative from positive tweets surrounding a brand, product, person or topic.

TERM

POSITIVE TWEETS

NEUTRAL TWEETS

NEGATIVE TWEETS

TOTAL TWEETS

st ives

70

384

11

465

15.05% POSITIVE



i really want to love st. ives apricot scrub, but it irritates my skin soo much :((view)



rt @kesiahosking: sunshine was smiling at you annie =)) rt @anniegreenwood st ives harbour basking in november sunshine http://flic.kr/p/8tk2sq (view)



sunshine was smiling at you annie =)) rt @anniegreenwood st ives harbour basking in november sunshine http://flic.kr/p/8tk2sq (view)



looking at st ives (uk:siv). great stats, but printing? (view)

82.58% NEUTRAL



@oldergirlbeauty GURL, I was all about the Aqua Net & the St. Ives liquid hairspray in the purple bottle. Where's my banana clip? (view)



RT @inscriptions: Loved the final episode of Junior Masterchef! Alexwill be at St Ives Village Sat 11th to show us a thing or two! (view)



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)



A Town On Canvas Called St lves http://ping.fm/onNWi (view).

2.37% NEGATIVE



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)



st ives face scrub receive negative comments. lots of it o.o (view)



@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)



sco prem: goal st ives city 2 towerhill blues 0 lucas k (43) (view)



sco prem: goal st ives city 1



Bonanza

Search

Try some Twitter trends: Romo Bonanza





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.



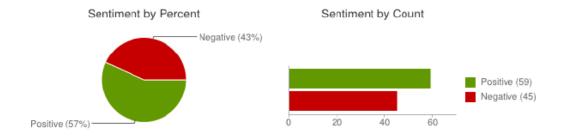
RT @ADReamGONe: RT @ADReamGONe: Man, I love Bonanza Imfao.

Sentiment140



microsoft English Search Save this search

Sentiment analysis for microsoft



Tweets about: microsoft

<u>Isaydumb</u>: <u>@Youporn</u>, in my humble opinion you have nothing to do on the <u>@Xbox</u> Live. What the fuck is <u>@Microsoft doing?!</u>
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 dont understand this disk haha. is it for microsoft project do you Posted 5 minutes ago

<u>jlebrech</u>: @rsslldnphy 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



Opinion Finder





Corpora News, debates, etc.



Annotation GATE, MPQA scheme



OpinionFinder

Version 1.x

Version 1.5

Version 1.4

Sample Annotations

Version 2.x

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 platformindependent.

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?

The task (1)

- Two main issues:
- What is an opinion? (to disinguish between opinions and facts in a text)
- How many opinions? One opinion is NOT sufficient (unless the opinion holder is very authoritative): we want to analyze many opinions, so we have also a problem of summarization

The Tasks (2)

Three types of opinion search:

- 1. Direct Opinions: sentiment expressions on some objects, e.g., products, events, topics, persons.
 - 1. E.g., "the picture quality of this camera is great"
 - 2. Subjective
- 2. Comparisons: relations expressing similarities or differences of more than one object. Usually expressing an ordering.
 - 1. E.g., "car x is cheaper than car y."
 - 2. 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

Task 1. Direct opinions:

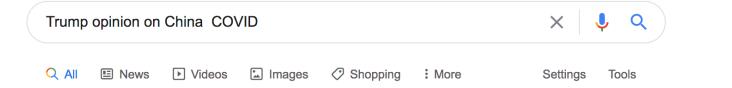
- Find the opinion of a person or organization (opinion holder) on a particular object or a feature of the object.
 - E.g., what is Bill Clinton's opinion on abortion?
- Find positive and/or negative opinions on a particular object (or some features of the object), e.g.,
 - customer opinions on a digital 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.
 - Trump's opinion on China and COVID

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.



About 732,000,000 results (0.52 seconds)

Top stories



Trump says China could have stopped Covid-19 and suggests US will seek damages

The Guardian

6 hours ago



China: US 'telling barefaced lies' on coronavirus - Live updates

Al Jazeera

1 hour ago



Trump says US may seek damages from China over Covid-19

RTE

11 hours ago

→ More for Trump opinion on China COVID

www.theguardian.com > world > apr > trump-says-china-could-have-s... ▼

Trump says China could have stopped Covid-19 and suggests ...

6 hours ago - President says US is conducting 'serious investigations' into Beijing's handling of virus.

www.theguardian.com > us-news > commentisfree > apr > trump-is-pl... ▼

Trump is playing a deadly game in deflecting Covid-19 blame ...

Apr 19, 2020 - As Mr 'Total Authority' keeps his focus firmly on re-election, he risks lives far beyond the United States.

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

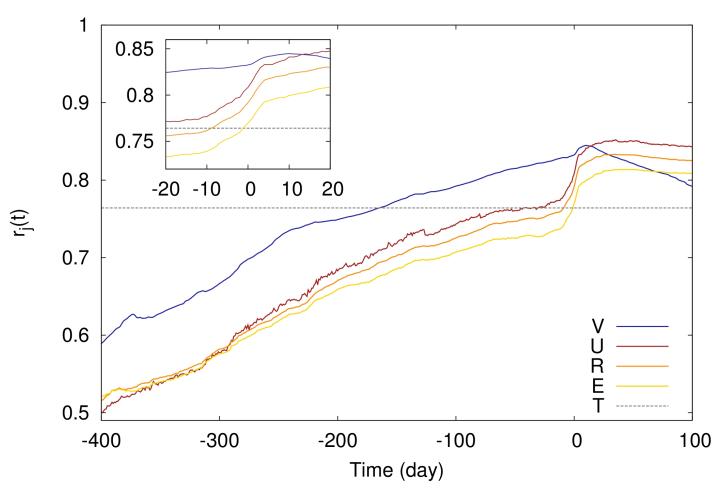
Task 2. Comparisons:

- 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"

Example of comparison result (Canon/Sony)



Task 3. Predictions



Temporal evolution of , the Pearson correlation of the box office revenue with different Predictors (e.g. based on Twitter or other blogs), for 24 movies.

Prediction: politics and sales

Politics

When Twitter Conversation Was Similar to Public Opinion

Supreme Court health care ruling (June 2012)

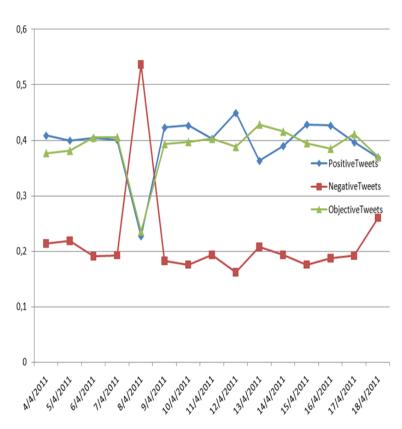
Public opinion	%	Twitter	%
Approve	36	Positive	52
Disapprove	40	Negative	48

Mitt Romney's choice of Paul Ryan (Aug 2012)

Public opinion	%	Twitter	%
Excellent/Good	28	Positive	28
Only fair/Poor	46	Negative	43
Don't know	26	Neutral	30

PEW RESEARCH CENTER Public opinion from Pew Research Center surveys; based on general public. See methodology for information on how Twitter reactions were derived.

Market



Monitoring of public opinion on Twitter for the keyword "milk".

Spike occurs on 8/4/2011 after a series of deaths in China relating to bad quality milk (source)

Other issues on opinion search: Ranking and result Visualization

Opinion ranking:

- should produce (at least) 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:

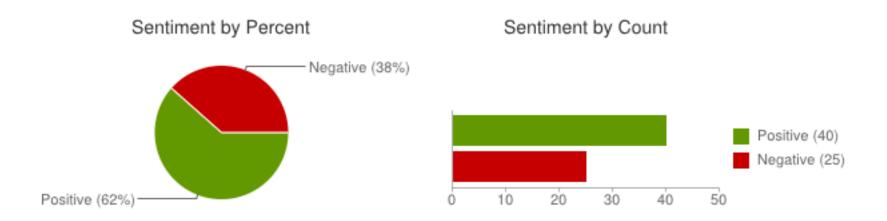
- Should the user reads all the top reviews? OR
- Should the system prepare a summary of the reviews?

Example of result visualization

Twitter Sentiment

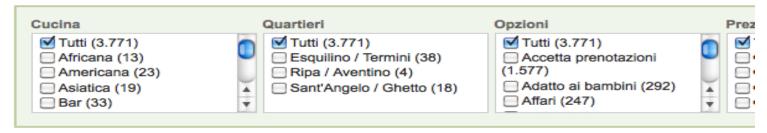


Sentiment analysis for Obama



Another example of visualization

Ristoranti: Roma



Recensioni attendibili

N. 2 di 3.389 ristoranti in Roma



Gli ospiti di he di...





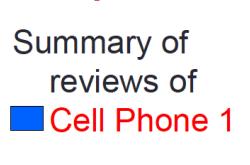


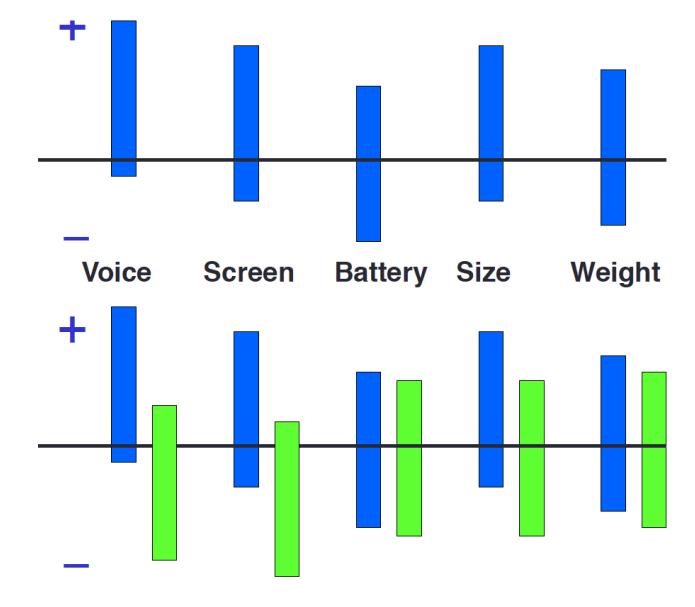
Tutti gli hotel in Ror



Twitter Sentiment Visualization







Comparison of reviews of

Cell Phone 1

Cell Phone 2



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.

Feature-Based Opinion Summary

(Hu & Liu, KDD-2004)

"I bought an iPhone a few days ago. It was such a nice phone. The touch screen was really cool. The voice quality was clear too. Although the battery life was not long, that is ok for me. However, my mother was mad with me as I did not tell her before I bought the phone. She also thought the phone was too expensive, and wanted me to return it to the shop. ..."

Feature Based Summary:

Feature1: Touch screen

Positive: 212

- The touch screen was really cool.
- The touch screen was so easy to use and can do amazing things.

. . .

Negative: 6

- The screen is easily scratched.
- I have a lot of difficulty in removing finger marks from the touch screen.

. . .

Feature2: battery life

. . .

Note: We omit opinion holders

. . . .

Opinion Mining Algorithms

- Machine learning
 - Naïve Bayes
 - Maximum Entropy Classifier
 - SVM

Assume pairwise independent features

- Markov Blanket Classifier (aka of Markov models)
 - Accounts for conditional feature dependencies
 - Allowed reduction of discriminating features from thousands of words to about 20
- All-flavors deep models (especially based on embeddings), see https://arxiv.org/pdf/1801.07883.pdf
- Unsupervised methods
 - Use opinion lexicons

Research Work	Document/Text Representation	Neural Networks Model	Use Attention Mechanism	Joint Modelling with Sentiment
Moraes et al. ³⁴	BoW	ANN (Artificial Neural Network)	No	-
Le and Mikolov ³⁵	Learning dense vector at sentence, paragraph, document level	Paragraph Vector	No	-
Glorot et al. ³⁶	BoW to dense document vector	SDA (Stacked Denoising Autoencoder)	No	Unsupervised data representation from target domains (in transfer learning settings)
Zhai and Zhang ³⁷	BoW to dense document vector	DAE (Denoising Autoencoder)	No	-
Johnson and Zhang ³⁸	BoW to dense document vector	BoW-CNN and Seq-CNN	No	-
Tang et al. ³⁹	Word embeddings to dense document vector	CNN/LSTM (to learn sentence representation) + GRU (to learn document representation)	No	-
Tang et al. ⁴⁰	Word embeddings to dense document vector	UPNN (User Product Neutral Network) based on CNN	No	User information and product information
Chen et al. ⁴¹	Word embeddings to dense document vector	UPA (User Product Attention) based on LSTM	Yes	User information and product Information
Dou ⁴²	Word embeddings to dense document vector	Memory Network	Yes	User information and product Information
Xu et al. ⁴³	Word embeddings to dense document vector	LSTM	No	-
Yang et al. ⁴⁴	Word embeddings to dense document vector	GRU-based sequence encoder	Hierarchical attention	-
Yin et al. ⁴⁵	Word embeddings to dense document vector	Input encoder and LSTM	Hierarchical attention	Aspect/target information
Zhou et al. ⁴⁶	Word embeddings to dense document vector	LSTM	Hierarchical attention	Cross-lingual information
Li et al. ⁴⁷	Word embeddings to dense document vector	Memory Network	Yes	Cross-domain information

Pros/Cons of ML approaches

Advantages:

- Tend to attain good predictive accuracy
 - Assuming you avoid the typical ML mishaps (e.g., over/under-fitting)

Disadvantages:

- Need for training corpus
 - Solution: automated extraction (e.g., Amazon reviews, Rotten Tomatoes) or crowdsourcing the annotation process (e.g., Mechanical Turk)

Domain sensitivity

- Trained models are well-fitted to particular product category (e.g., electronics) but underperform if applied to other categories (e.g., movies)
- Solution: train a lot of domain-specific models or apply domain-adaptation techniques
- Particularly for Opinion Retrieval, you'll also need to identify the domain of the query!
- Often difficult/impossible to rationalise prediction output

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

Lexicon-based solutions

- Detect/extract the polarity of opinions, based on sentiment dictionaries
- Word-lists where each token is annotated with an 'emotional' value
 - e.g., positive/negative words or words that express anger, fear, happiness, etc.
- Add syntactic and prose rules to estimate the overall polarity of text:
 - Negation detection: "the movie wasn't good"
 - Exclamation detection: "great show!!"
 - Emoticon detection: "went to the movies ©"
 - Emphasis detection: "You are goooood"
 - Intensifier, diminisher word detection: "Very good movie" vs. "good movie"

(Basic) lexicon-based approach

- Detect emotion in two independent dimensions:
 - Positive: D_{pos}: {1, 2,... 5}
 - Negative: D_{neg}: {-5, -4,... -1}
- (optional) Predict overall polarity by comparing them:
 - If $D_{pos} > |D_{neg}|$ then positive
- Example: "He is brilliant but boring"
 - Emotion('brilliant')=+3

$$D_{pos} = +3$$
, $D_{neg} = -2 => positive$

- Emotion('boring')=-2
- Negation detection: "He isn't brilliant and he is boring"
 - Emotion(NOT 'brilliant') = -2

$$D_{pos} = +1$$
 (default),

Decreased by 1 and sign reversed

$$D_{neg}$$
=-3 => negative

- Exclamation detection: "He is brilliant but boring!!"
 - Might increase polarity (either neg or pos)

Pros/Cons of the approach

Advantages:

- Can be fairly accurate, independent of environment
- No need for training corpus
- Can be easily extended to new domains with additional affective words
 - e.g., "amazeballs"
- Can be easy to rationalise prediction output
- More often used in Opinion Retrieval

Disadvantages:

- Compared to a well-trained, in-domain ML model they typically underperform
- Sensitive to affective dictionary coverage

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

Summary

- Sentiment analysis is a difficult task
- The difficulty increases with the nuance and complexity of opinions expressed
- Product reviews, etc are relatively easy
- Books, movies, art, music are more difficult
- Policy discussions, indirect expressions of opinion more difficult still
- Non-binary sentiment (political leanings etc) is extremely difficult
- Patterns of alliance and opposition between individuals become central (e.g.,

https://www.cs.virginia.edu/~hw5x/paper/KDD2018-MMB.pdf)

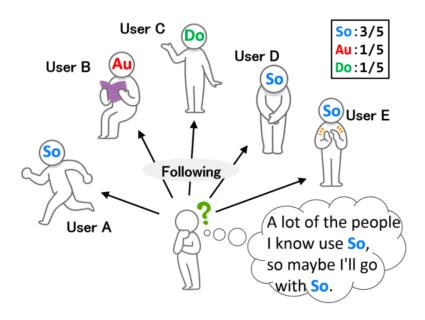
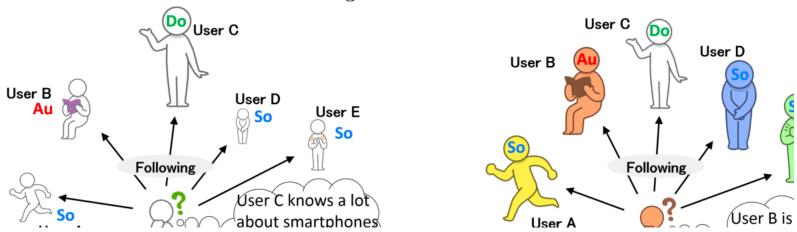


Figure 1: Basic Voter Model



https://www.semanticscholar.org/paper/Modeling-Opinion-Formation-by-Incorporating-Users%27-Fushimi-Ono/2417d76b3f29d3e75c2718981060122b8b98c1be/figure/0