

Sentiment Analysis in Practice

Yongzheng (Tiger) Zhang

Dan Shen*

Catherine Baudin

December 12, 2011@ICDM'11

* dianping.com



Outline

1. Introduction
2. Sentiment Identification & Classification
3. Key Applications
4. Examples
5. Conclusions

Roadmap

1. Introduction
2. Sentiment Identification & Classification
3. Key Applications
4. Examples
5. Conclusions

Example 1 – A Wikipedia Entry



WIKIPEDIA
The Free Encyclopedia

- Main page
- Contents
- Featured content
- Current events
- Random article
- Donate to Wikipedia

- Interaction
 - Help
 - About Wikipedia
 - Community portal
 - Recent changes
 - Contact Wikipedia

- Toolbox
- Print/export

- Languages
 - العربية
 - Български
 - Català
 - Česky
 - Deutsch

Log in / create account

Article Discussion

Read

Edit

View history

Search



Machine learning

From Wikipedia, the free encyclopedia

For the journal, see [Machine Learning \(journal\)](#).



This article **is missing citations or needs footnotes**. Please help add inline citations to guard against copyright violations and factual inaccuracies. *(August 2008)*

Machine learning, a branch of artificial intelligence, is a scientific discipline concerned with the design and development of algorithms that allow computers to evolve behaviors based on empirical data, such as from sensor data or databases. A learner can take advantage of examples (data) to capture characteristics of interest of their unknown underlying probability distribution. Data can be seen as examples that illustrate relations between observed variables. A major focus of machine learning research is to automatically learn to recognize complex patterns and make intelligent decisions based on data; the difficulty lies in the fact that the set of all possible behaviors given all possible inputs is too large to be covered by the set of observed examples (training data). Hence the learner must generalize from the given examples, so as to be able to produce a useful output in new cases. Machine learning, like all subjects in artificial intelligence, requires cross-disciplinary proficiency in several areas, such as probability theory, statistics, pattern recognition, cognitive science, data mining, adaptive control, computational neuroscience and theoretical computer science.

Contents [hide]

- Definition
- Generalization
- Human interaction
- Algorithm types
- Theory
- Approaches
 - Decision tree learning

Example 2 – A Milo Product Review

The screenshot displays the Milo local shopping interface. At the top left is the Milo logo with a dog icon and the text "Milo local shopping". To the right is a search bar with the text "Find in-stock products near you" and a search button. Below the search bar is a navigation menu with categories: "Browse Categories", "Electronics", "Computers & Accessories", and "Tablets".

The main product listing is for an "Apple® - iPad^{2nd} with Wi-Fi + 3G - 64GB" priced at "\$679.00". It features a 5-star rating from 16 reviews, a "Like" button, and a "Sign Up to see what your friends like" link. A "Set Price Alert" button is also present.

Under the heading "Availability near Youngstown", a Best Buy store is listed for \$679.00. A "View on Store Website" button is provided. A list of store locations follows:

- IN STOCK San Jose
- IN STOCK San Jose
- IN STOCK Sunnyvale
- IN STOCK Mountain View
- OUT OF STOCK San Jose

A link for "1 more location" is also shown. Below the list is a disclaimer: "Milo cannot guarantee the accuracy of pricing, product, availability and other information provided by third parties. See Terms for more details." A map of the San Jose area is displayed, showing several Best Buy locations marked with 'B' icons.

On the left side of the product listing, there is a large image of the iPad. Below the image are the following review details:

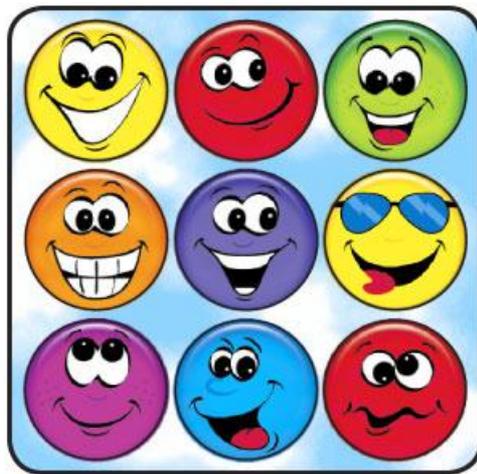
- 16 Reviews**
- 5 star: (14)
- 4 star: (1)
- 3 star: (1)
- 2 star: (0)
- 1 star: (0)

A "Most Helpful" dropdown menu is set to "Most Helpful". Below the review statistics are two review snippets:

- Love this device** (5 stars): "Great device for accessing the Internet and doing anything Apple related. The ..."
- Awesome!** (5 stars): "I was the hero of the trip recently on a trip to Chicago(actually writing this ..."

Facts vs. Opinions

- **Facts:** **objective** expressions about entities, events and their attributes, e.g. “*I bought an iPhone yesterday*”
- **Opinions:** **subjective** expressions of sentiments, attitudes, emotions, appraisals or feelings toward entities, events and their attributes, e.g. “*I really love this new camera*”



Facts vs. Opinions – An Article from NYT

- http://www.nytimes.com/2009/08/24/technology/internet/24emotion.html?_r=2&pagewanted=all

Mining the Web for Feelings, Not Facts

By ALEX WRIGHT
Published: August 23, 2009

Computers may be good at crunching numbers, but can they crunch feelings?

 Enlarge This Image



Peter DaSilva for The New York Times
Margaret Francis and Mars Hall of the San Francisco company Scout Labs.

 Enlarge This Image



Minh Uong/The New York Times

 Enlarge This Image



The rise of blogs and social networks has fueled a bull market in personal opinion: reviews, ratings, recommendations and other forms of online expression. For computer scientists, this fast-growing mountain of data is opening a tantalizing window onto the collective consciousness of Internet users.

An emerging field known as sentiment analysis is taking shape around one of the computer world's unexplored frontiers: translating the vagaries of human emotion into hard data.

This is more than just an interesting programming exercise. For many businesses, online opinion has turned into a kind of virtual currency that can make or break a product in the marketplace.

Yet many companies struggle to make sense of the caterwaul of complaints and compliments that now swirl around their products online. As sentiment analysis tools begin to take shape, they could not only help businesses improve their bottom lines, but also eventually transform the experience of searching for information online.

Several new sentiment analysis companies are trying to tap into the growing business interest in what is being said online.

"Social media used to be this cute project for 25-year-old consultants," said Margaret Francis, vice president for product at Scout Labs in San Francisco. Now, she said, top executives "are recognizing it as an incredibly rich vein of market intelligence."

 TWITTER

 SIGN IN TO E-MAIL

 PRINT

 REPRINTS

 SHARE

WIN WIN
NOW PLAYING

Some Exceptions!

- Not all subjective sentences contain opinions, e.g.
 - “*I want a phone with good voice quality*”
- Not all objective sentences contain no opinions, e.g.
 - “*The earphone broke in just two days!*”

History of Sentiment Analysis

- Individuals relied on a circle of
 - Family
 - Friends
 - Colleagues
- Organizations used
 - Polls
 - Surveys
 - Focus groups
- Sentiment analysis is growing rapidly



Why Now?

- WWW has facilitated *user-generated content*

- e-commerce sites
- Forums
- Discussion groups
- Blogs
- Twitter, etc.

The image displays two screenshots of e-commerce websites. The left screenshot shows an eBay discussion board for the 'iPhone Selling app feedback' category. It features a 'Spring Seller Update' announcement and several user posts with replies, including one from 'ganga.developers...' dated Mar 11, 2011, and another from 'Indelbesh (OT)' dated Apr 1, 2010. The right screenshot shows an Epinions.com product page for a Canon PowerShot camera. It includes a search bar, a product image, a 'General Rating' of 4.5 stars, a 'Lowest Price' of \$111.31, and a list of 19 stores with their respective prices and shipping options. A 'What's HOT in Video' banner for 'Ever Want to Star in the Movies?' is also visible.

- Advances in NLP and text-processing
- Distributed computing: e.g. Hadoop, Cloud
- Our focus: automatically mining opinions, challenging but useful

Opinions and Market Research

Market Research 3.0 Is Here: Attitudes Meet Algorithms in Sentiment Analysis

BY KEVIN RANDALL Fri Sep 18, 2009

This is the marketer's and researcher's dream.

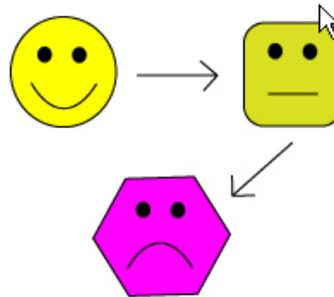
Reconciling the natural tensions that challenge and befuddle brand planning:

- Feelings & Facts
- Sentiments & Statistics
- Qualitative & Quantitative
- Focus Groups & Surveys
- Subjective & Objective
- Why & What
- Art & Science

I'll admit, when I first heard about Google, Facebook, and Nielsen studying, decoding and monitoring language and chatter on the Web and "listening to conversations," the consumer part of me got a little bit of the creeps (Big Brother idea).

On the other hand, the market researcher part of me was excited about all of the possibilities. Market research has been stale for a while. Everyone knows about the limitations of the traditional focus group and survey. Do group respondents even tell the truth in an artificial setting where they are served finger sandwiches and paid \$100? How can the group think be weeded out to get a real picture of the market? Are the right people answering online surveys? Are panelists professional survey respondents or representative customers?

The explosion of social media channels has the potential to revolutionize market research. New social media-based studies can be conducted more cheaply and efficiently, in real-time and may more accurately capture individual and group opinions. Companies are already



<http://www.fastcompany.com/blog/kevin-randall/integrated-branding/market-research-30-here-attitudes-meet-algorithms-sentiment-a?partner=rss> is from a market-researcher's perspective

Practical Applications – Business & Organizations

- Brand analysis
- Marketing
- Customer voice: e.g. products, tourism
- Event monitoring, e.g. site outage
- Commercial examples
 - [Radian6](#)
 - [Lexalytics](#)

Practical Applications – Individuals

- Shopping research, e.g. product reviews

- http://www1.epinions.com/prices/Canon_PowerShot_SD1300_IS_IXUS_105_Digital_Camera

Epinions.com
a shopping.com company

Join Epinions | Learn More | Sign In

CARS BOOKS MOVIES MUSIC COMPUTERS & SOFTWARE ELECTRONICS GIFTS HOME & GARDEN KIDS & FAMILY OFFICE SUPPLY SPORTS TRAVEL MORE...

Search for Digital Cameras Search

Home > Electronics > Digital Cameras > Canon PowerShot Powershot SD1300 IS / Digital IXUS 105 Digital Camera

Canon PowerShot Powershot SD1300 IS / Digital IXUS 105 Digital Camera

Overall Rating: ★★★★★
23 consumer reviews

Lowest Price: \$111.31
at [BuyDig.com](#) **SEE IT**

[Subscribe to reviews on this product](#)

Compare Prices Read Reviews (23) View Details Write a Review

1-19 of 19 stores [Check Tax & Shipping](#)

	Sort by Store Rating	Sort by Base Price
<p>Canon Powershot Sd1300 Is Digital Elph 12.1 Megapixe... Earn 2% eBay Bucks on qualifyi... (In stock) Canon PowerShot SD1300 IS Digital ELPH - Digital camera - compact - 12.1 Mpix - optical zoom: 4 x - supported memory: MMC, SD, SDXC, SDHC, MMCplus - g... More at eBay Read 1198 customer reviews on eBay</p>	<p>Featured Store Trusted Store 1198 store ratings</p>	\$129.99 SEE IT
<p>Canon 12.1 Megapixel PowerShot; SD1300 IS Digital ELPH; Camera- Choose Colo... (In stock) It's not just for still images anymore. The Canon; 12.1 megapixel PowerShot; SD1300 IS Digital ELPH; camera can record high-definition video in crisp,... More at Sports Authority Read 78 customer reviews on Sports Authority</p>	<p>Featured Store Trusted Store 78 store ratings</p>	\$239.99 SEE IT
<p>Canon SD1300 Digital Camera- Pink 12MP, 4X, 2.7", 28mm Wide 30-Day Money-back Guarantee! (In stock) 12MP, 4X, 2.7", 28mm Wide 2-11/16" LCD screen,optical image stabilization,wide-angle capability ideal for sweeping landscapes More at Crutchfield Read 4634 customer reviews on Crutchfield</p>	<p>Featured Store Trusted Store 4634 store ratings</p>	\$179.99 FREE SHIPPING! SEE IT

Practical Applications – Individuals (cont.)

- Seeking out opinions on topics such as finance

- <http://community.nytimes.com/comments/bucks.blogs.nytimes.com/2011/03/10/where-all-the-mortgage-documents-go/?scp=2&sq=refinance%20a%20loan&st=cse>

Where All the Mortgage Documents Go

Chase now scans all mortgage-related documents that come in via fax mail.

Share your thoughts.

[Back to Blog Post »](#)

13 Readers' Comments

[Post a Comment »](#)

[ALL COMMENTS](#)

[HIGHLIGHTS](#)

[READERS' RECOMMENDATIONS](#)

[REPLIES](#)

Oldest Newest

1. Doug Rife
Sarasota, FL
March 11th, 2011
8:09 am

There was time not too long ago when banks were regarded as having superior recording keeping technologies far beyond ordinary businesses. Banks were among the first businesses to embrace computers long before there were any PCs, scanners or laser printers. But now we see the myth of banks being meticulous bookkeepers, at least the for megabanks in their mortgage departments. It's shocking to learn that Chase just started scanning mortgage documents in 2009. All banks should have been scanning and routinely backing up all mortgage documents long before the financial crisis hit. Cheap scanners have been around for ten years or more. Now, it turns out that many banks can't even keep straight which mortgages they own, which ones are current and which are delinquent. Incompetence of the highest order.

[Recommend](#)

Recommended by 10 Readers

2. Rolando
New York, NY
March 11th, 2011
8:09 am

When I applied last year for a Citi refi, they sent an encrypted PDF that I could sign electronically and send back to them immediately; much more efficient!

[Recommend](#)

Recommended by 0 Readers

Practical Applications – Advertising

- Placing ads in user-generated content

- <http://www.zurich-hotels-booker.com/why-get-a-canon-powershot-a1200-98757.html>

The screenshot shows a blog post on a website. The main content area has a title "Why Get a Canon Powershot A1200" in a dark red font. Below the title, there is a date "22 Mar 2011 at 06:09" and a byline "admin". The post text discusses the benefits of the Canon Powershot A1200 camera for hiking and travel. A sidebar on the right contains two sections: "Categories" with a list of topics like "Arts & Entertainment", "Business & Finance", and "Education"; and "Blogroll" with links for "Feed", "Log in", and "WordPress". At the bottom of the post, there is a "Share / Save" button with social media icons for Facebook, Twitter, and LinkedIn.

Practical Applications – Opinion Search

- Providing general search for opinions, not facts
- Tons of social media applications

- [TED](#)
- [TWECAN](#)
- [SocialMention](#)
- [TweetSentiments](#)

socialmention*

Blogs Microblogs Bookmarks Comments Events Images News Video Audio Q&A Networks All [Advanced Search Preferences](#)

Japan Nuclear crisis

37% strength	1:1 sentiment
21% passion	39% reach
16 seconds avg. per mention	
last mention 45 seconds ago	
407 unique authors	
62 retweets	

Sentiment

positive	65
neutral	567
negative	66

Top Keywords

nuclear	975
japan	738
plant	232
fukushima	204

Mentions about Japan Nuclear crisis

Sort By: Results: Results 1 - 15 of 698 mentions.

- RT @reuters: France and U.S. to help Japan in nuclear crisis <http://reut.rs/hPfEg0>
twitter.com/canaXD/statuses/52845499623018496
45 seconds ago - by @canaXD on [twitter](#)
- RT @Reuters France and U.S. to help Japan in nuclear crisis <http://reut.rs/hPfEg0>
twitter.com/TroubledDays/statuses/52845490731094016
47 seconds ago - by [troubleddays](#) on [stumbleupon](#)
- Farmers' Livelihoods Wither in Japan's Nuclear Crisis - New York Times: <http://bit.ly/e9wNyd>
twitter.com/EDUBEAT/statuses/52845480836730880
49 seconds ago - by [edubeat](#) on [stumbleupon](#)
- RT @Reuters: France and U.S. to help Japan in nuclear crisis <http://reut.rs/hPfEg0>
twitter.com/Sugib/statuses/52845448708362241
57 seconds ago - by [sugi](#) on [stumbleupon](#)

Opinion Search – One More Example

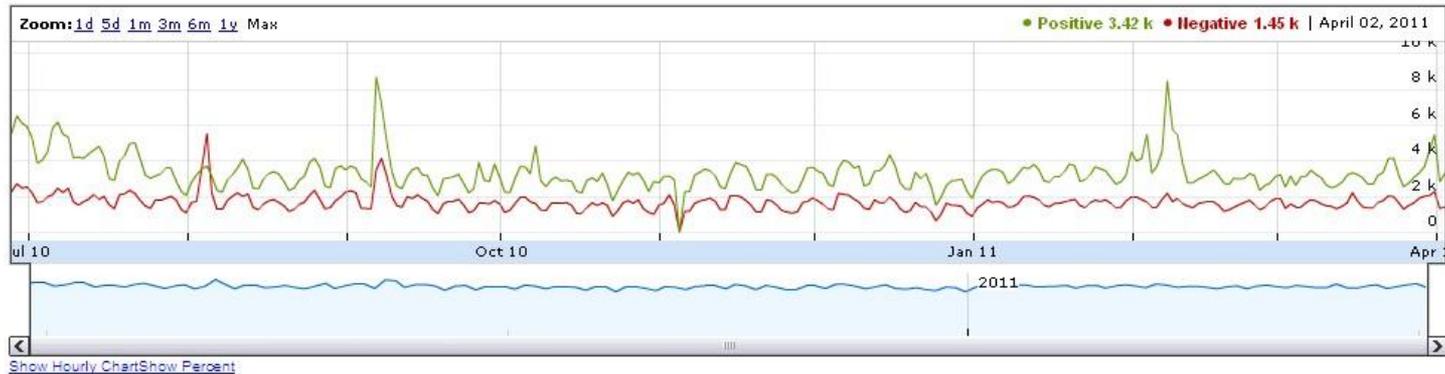
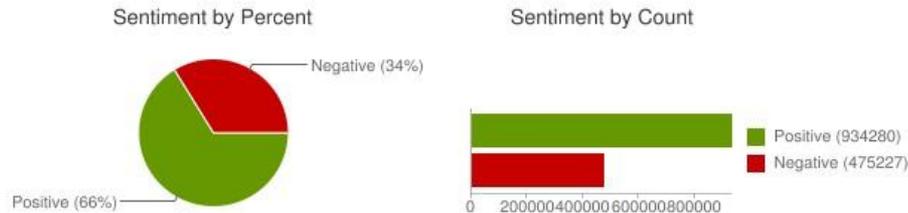
[Sign in](#) | [Help](#)

Twitter Sentiment

Type in a word and we'll highlight the good and the bad

[Save this search](#)

Sentiment analysis for google



Tweets about: google

Choose a date range: to (Format: dd/mm/yyyy. You can also adjust tr) The results for this query are:

The Problem

- A review example: (1) *I bought an iPhone a few days ago.* (2) *It was such a nice phone.* (3) *The touch screen was really cool.* (4) *The voice quality was clear too.* (5) *Although the battery life was not long, that is ok for me.* (6) *However, my mother was mad with me as I did not tell her before I bought it.* (7) *She also thought the phone was too expensive, and wanted me to return it to the shop. . . .*
- Facts vs. opinions
- Opinions have *targets* (objects and their attributes) on which opinions are expressed

Definitions

- **Object**: an entity that can be a product, service, individual, organization, event, or topic, e.g. iPhone
- **Attribute**: an object usually has two types of attributes
 - Components, e.g. battery, keypad/touch screen
 - Properties, e.g. size, weight, color, voice quality
- **Explicit and implicit attributes**
 - Explicit attributes: those appearing in the opinion, e.g. “*the battery life of this phone is too short*”
 - Implicit attributes: those not appearing in the opinion, e.g. “*this phone is too large*” (on attribute *size*)
- **Opinion holder**: the person or organization that expresses the opinion
- **Opinion orientation (polarity)**: positive, negative, or neutral
- **Opinion strength**: level/scale/intensity of opinion indicating how strong it is, e.g. contented → happy → joyous → ecstatic

Key Elements of an Opinion

- *Opinion*: a person or organization that expresses a positive or negative sentiment on a particular attribute of an object at a certain time
- *Quintuple*: *<object, attribute, orientation, opinion holder, time>*
 - Some information may be implied due to pronouns, context, or language conventions
 - Some information available from document attributes
 - In practice, not all five elements are needed

Direct vs. Comparative Opinions

- *Direct opinion*: sentiment expressions on one or more attributes of an object, e.g. products, services, events
 - “*The voice quality of this phone is fantastic*”
 - “*After taking this medicine, my left knee feels worse*”
- *Comparative opinion*: relations expressing similarities or differences between two or more objects based on some of the shared attributes of the objects, e.g.
 - “*The voice quality of camera x is better than that of camera y*”
- There are some difficult cases which are not covered, e.g. “*The view finder and the lens of the camera are too close to each other*”

Explicit vs. Implicit Opinions

- *Sentence subjectivity*: an objective sentence expresses some **factual** information about the world, while a subjective sentence expresses some **personal** feelings or beliefs
- *Explicit opinion*: an opinion on an attribute explicitly expressed in a subjective sentence, e.g.
 - “*The voice quality of this phone is amazing*”
 - “*This camera is too heavy*”
- *Implicit opinion*: an opinion on an attribute implied in an objective sentence, e.g.
 - “*The headset broke in two days*”
 - “*Please bring back the old search*”

Outline

1. Introduction
2. **Sentiment Identification & Classification**
3. Key Applications
4. Examples
5. Conclusions

A Product Review Example

- An epinions.com product review for [Canon PowerShot SD1300 IS / Digital IXUS 105 Digital Camera](#)

“My Canon Powershot takes *great* pictures! ... My friend had gotten one about a year ago and *she loves it*. So, after seeing her enthusiasm about it I decided to get one and *I will never go back to any other camera*. I *absolutely love* this camera. I believe that *every person on Earth should own one of these*. ... It is *amazing!* ... *There is not one thing I hate about this product*, which is strange because I am a very picky person! ...”
- What do we see in this example?

Sentiment Analysis Tasks

- Goal: identify and classify opinions
- Task 1. *Sentiment identification* (Subjectivity identification): identify whether a piece of text expresses opinions
- Task 2. *Sentiment orientation classification*: determine the orientation of an opinionated text

Sentiment Analysis Levels

- *Document-level*: identify if the **document** (e.g. product reviews, blogs, forum posts) expresses opinions and whether the opinions are positive, negative, or neutral
- *Sentence-level*: identify if a **sentence** is opinionated and whether the opinion is positive, negative, or neutral
- *Attribute-level*: extract the **object attributes** (e.g. image quality, zoom size) that are the subject of an opinion and the opinion orientations
- As the object becomes more granular, the intensity/difficulty increases

Document-level Sentiment Analysis

- Tasks: identify if the document expresses opinions and if yes classify the document into positive, negative, or neutral based on the overall sentiments expressed by opinion holders
- Assumptions:
 - the document is opinionated on a single object
 - the opinions are from a single opinion holder
- Similar to but different from topic-based text classification
 - In topic-based text classification, topic words are important
 - In sentiment classification, *opinion words* are more important, e.g. wonderful, fabulous, terrible

Opinion Words

- Also known as *polarity words*, *sentiment words*, *opinion lexicon*, or *opinion-bearing words*, e.g.
 - Positive: wonderful, elegant, amazing
 - Negative: horrible, disgusting, poor
- *Base* type (examples above) and *comparative* type (e.g. better, worse)
- How to generate them: more on this later

A Simple Method – Counting Opinion Words

- Opinion/polarity words: dominating indicators of sentiments, especially **adjectives**, **adverbs**, and **verbs**, e.g. “I absolutely *love* this camera. It is *amazing!*”.
- Pre-defined opinion words: good, terrible, ... (more on this later)
- Assign *orientation score* (+1, -1) to all words
 - Positive opinion words (+1): *great, amazing, love*
 - Negative opinion words (-1): *horrible, hate*
 - Strength value [0, 1] can be used too
- The orientation score of the document is the sum of orientation scores of all opinion words found
 - The previous review has an orientation score of $4 - 1 = 3$

Rule-based Method

- Is simply counting opinion words good enough? No!
 - “There is *not* one thing I *hate* about this product” → Wrong
- We need to handle negation: “*not* ... *hate*” implies *like*
 - Simple rules can be manually created
 - “not ... *negative*” → *positive*
 - “never ... *negative*” → *positive*
 - The previous review has a score of $4 + 1 = 5$
 - Note: negation needs to be handled with care, e.g. “not” in “not only ... but also” does not change the orientation

Terminology

- *Pattern/rule*: a sequence of tokens
- *Token*: an abstraction of a word, represented using lemma of a word, polarity tag, or Part Of Speech (POS) tag. Two special tokens:
 - TOPIC: an attribute, e.g. size, weight
 - GAP_digit_digit: how many words can be skipping between two tokens to allow more tolerant matching, e.g. GAP_1_2
- *Polarity tag*: positive, negative, neutral, NOT (negation)
- *POS tag*: NN (noun), VB (verb), JJ (adjective), RB (adverb), IN (preposition) ...

Basic Opinion Rules – Label Sequential Pattern (LSP) Matching

- Subject {like | adore | want | work} TOPIC → positive, e.g.
 - “I like the old camera”
- Subject {is | are} {great | fantastic | simple | easy} → positive, e.g.
 - “This camera is fantastic”
- TOPIC GAP_0_3 NOT work → negative, e.g.
 - “*The new search still does not work*”
- Please do NOT VB → negative, e.g.
 - “*Please do not roll out this new search!*”
- NOT GAP_0_3 {want | think | believe | need | get} → negative, e.g.
 - “*I do not want large size pictures in the Gallery window*”
- {get | bring | give | put | change} GAP_0_3 TOPIC GAP_0_3 back → positive
 - “Please put the old search and browse back!”

Limitation of Rule-based System

- An expensive task:
 - Only a limited number of opinion words can be found
 - Only a limited number of patterns can be created
- Can we automate the task with limited manual work?
e.g. find opinion words and their orientations automatically

Automatically Finding Opinion Words [Turney, ACL2002]

- Data: reviews from epinions.com on automobiles, banks, movies and travel destinations
- Step 1. perform part of speech (POS) tagging and extract phrases containing adjectives and adverbs based on manually specified patterns

Table 1. Patterns of POS tags for extracting two-word phrases

First word	Second word	Third word (Not Extracted)
1. JJ	NN or NNS	anything
2. RB, RBR, or RBS	JJ	not NN nor NNS
3. JJ	JJ	not NN nor NNS
4. NN or NNS	JJ	not NN nor NNS
5. RB, RBR, or RBS	VB, VBD, VBN, or VBG	anything

- e.g. extract two consecutive words if the first word is adjective, the second is a noun and the third (which is not extracted) is anything: “this camera produces beautiful pictures”

Pointwise Mutual Information (PMI)

- Step 2. estimate the orientation of each extracted phrase using the PMI measure
 - PMI is the amount of information that we acquire about the presence of one of the words when we observe the other

$$PMI(term_1, term_2) = \log_2 \left(\frac{\Pr(term_1 \wedge term_2)}{\Pr(term_1) \Pr(term_2)} \right).$$

- The opinion orientation (OO) of a phrase is computed based on its association with the positive reference word “excellent” and its association with the negative reference word “poor”

$$oo(phrase) = PMI(phrase, “excellent”) - PMI(phrase, “poor”).$$

PMI (cont.)

- Estimate probabilities with number of hits of search query
 - For each search query, search engine returns the number of relevant documents to the query, which is the number of hits
- Turney used AltaVista which had a NEAR operator, which constrains the search to documents that contain the words within 10 words of one another, in either order

$$oo(\textit{phrase}) = \log_2 \left(\frac{\textit{hits}(\textit{phrase NEAR "excellent"})\textit{hits}(\textit{"poor"})}{\textit{hits}(\textit{phrase NEAR "poor"})\textit{hits}(\textit{"excellent"})} \right)$$

- Examples:
 - “*low fees*”, “JJ NNS”, 0.333
 - “*unethical practices*”, “JJ NNS”, -8.484
 - “*low funds*”, “JJ NNS”, -6.843

Sentiment Orientation Classification

- Step 3. Compute the average semantic orientation of all phrases in the review
 - Classify as positive (recommended) or negative (not recommended) based on the sign of the average
 - Final classification accuracy:
 - Automobiles – 84%
 - Banks – 80%
 - Movies – 66%
 - Travel destinations – 71%

Note: Recent variations use more than two words to determine the orientation

Polarity Words Generation

- Manual: effective but expensive
- Dictionary-based: use a seed list and grow the list, e.g.
 - [SentiWordNet](#)
- Corpus-based: rely on *syntactic* or *co-occurrence patterns* in large text corpora [Hazivassiloglou & McKeown, ACL1997; Turney, ACL2002; Yu & Hazivassiloglou, EMNLP2003, Kanayama & Nasukawa, EMNLP2006; Ding & Liu, SIGIR2007]

ANEW: Affective Norms for English Words

<http://csea.php.ufl.edu/media/anewmessage.html> is specific to English

NIMH CENTER FOR THE STUDY OF EMOTION AND ATTENTION University of Florida

[Home](#)

[About Us](#)

[Center Sites/People](#)

[Research Projects](#)
Davis, Lang, Bradley, Patrick, Öhman,
Cacioppo

[Science Technology Core](#)

[Media Core](#)
IAPS, IADS, ANEW, ANET

[Research Training &
Administration Core](#)

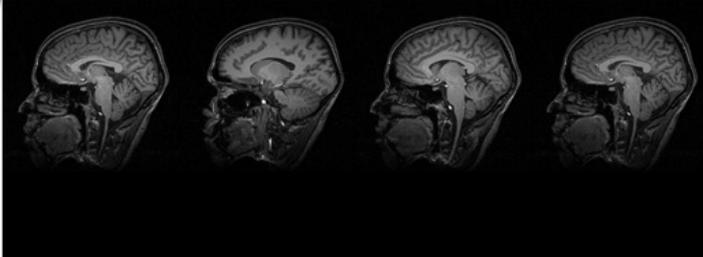
[Publications](#)

[Job Openings](#)

[Useful Links](#)

[External Advisory Board](#)

[Contact Information](#)



ANEW Message

Dear Colleague

Thank you for your interest in the Affective Norms for English Words (ANEW). The ANEW provides a set of normative emotional ratings for a large number of words in the English language. This set of verbal materials have been rated in terms of pleasure, arousal, and dominance in order to create a standard for use in studies of emotion and attention.

Considering our exclusively scholarly interests and our status as a unit of the State University, we do not sell these materials. We provide them solely for use in academic, not-for-profit research at recognized educational institutions.

To receive the ANEW for your research, please fill out and submit the ANEW Researcher Information Form. Upon submission of this form, you will receive an email asking you to print, sign and fax the ANEW User Agreement. Within 30 days of receiving your request, you will receive a username and password, and the link to download the ANEW. This link should not be shared, distributed, or linked to from another website.

Because part of the funded mission of CSEA is to provide materials to research labs, it is helpful for us to be able to list the laboratories benefiting from this work. Thus, we would like to take this opportunity to ask that you refer other researchers who would like to receive these materials to us, rather than distributing the ANEW to others yourself.

Thanks again for your interest.

Margaret M. Bradley, Ph.D.
Coordinator CSEA Media Core

Peter J. Lang, Ph.D.
CSEA Director

Dictionary-based Approach – Bootstrapping

- Step 1. manually collect a small set of opinion words with known orientations
 - {*glad*}
- Step 2. search an online dictionary (e.g. [WordNet](#) [FellbaumC1998]) for their synonyms and antonyms to grow the word set [Hu & Liu, KDD2004; Kim & Hovy, COLING2004]
 - {*glad, happy, joyful, delighted*}; {*sad, unhappy, sorry, heart-broken*}
- Repeat steps 1 & 2 until no more new words are found
- Finally, manual inspection may be done for correction
- Additional information (e.g. glosses) from WordNet [Andreevskaia & Bergler, EACL2006] can be used

Limitations of Dictionary-based Approach

- Cannot identify *context-dependent* opinion words
- Example 1: small
 - “the LCD screen is too **small**” vs.
 - “the camera is very **small** and easy to carry”
- Example 2: long
 - “it takes a **long** time to focus” vs.
 - “the battery life is **long**”

Corpus-based Approach

- Rely on *syntactic* or *co-occurrence patterns* in large text corpora [Hazivassiloglou & McKeown, ACL1997; Turney, ACL2002; Yu & Hazivassiloglou, EMNLP2003, Kanayama & Nasukawa, EMNLP2006; Ding & Liu, SIGIR2007]
- Can find *domain dependent* orientations and/or *context dependent* ones!

An Example – Predicting the Sentiment Orientations of Adjectives

[Hazivassiloglou & McKeown, ACL1997]

- Start with a list of seed opinion adjective words
- Use *linguistic constraints* on connectives to identify additional adjective opinion words and their orientations
 - *Sentiment consistency*: conjoined adjectives usually have the same orientations → This car is beautiful *and* spacious. (if beautiful is positive, then spacious is positive too)
 - Rules can be designed for different connectives: AND, OR, BUT, EITHER-OR, NEITHER-NOR
- Use *log-linear model* to determine if two conjoined adjectives are of the same or different orientations
- Use *clustering* to produce two sets of words, i.e. positive and negative
- Data corpus: 21 million words from 1987 Wall Street Journal

Another Example – Handling Context Dependency

[Ding & Liu, 2007, Lu et al, WWW2011]

- Finding domain opinion words is NOT sufficient
 - one word may indicate different opinions in the same domain, e.g. “The battery life is long” vs. “It takes a long time to focus”.
- *Opinion context* idea: use pairs of (*object_attribute*, *opinion_word*)
- Determining opinion words and their orientations together with the object attributes
- It can be used without a large corpus

Manual → Automated: Can We Learn from Examples?

Canon PowerShot Powershot SD1300 IS / Digital IXUS 105 Digital Camera

Overall Rating: ★★★★★
[23 consumer reviews](#)

Lowest Price: \$111.31

Ads by Google
[Canon Powershot SD1300 IS](#)
 Free Shipping Canon Powershot SD1300 IS 12MP Digital ELPH.

Compare Prices

Great Camer
 Written: Feb 06 11

Product Rating: ★★★★★
[23 consumer reviews](#)

Lowest Price: \$111.31

Ads by Google
[Canon Powershot SD1300 IS](#)
 Free Shipping Canon Powershot SD1300 IS 12MP Digital ELPH.
[www.buydig.com/Canon_Free_Shipping](#)

Great Camera
 Written: Feb 06 11

Product Rating: ★★★★★
[23 consumer reviews](#)

Lowest Price: \$111.31
 at [BuyDig.com](#) **SEE IT!**

Subscribe to reviews on this product

Ads by Google
[Canon Powershot SD1300 IS](#)
 Free Shipping Canon Powershot SD1300 IS 12MP Digital ELPH.
[www.versadeals.com](#)

Compare Prices

Read Reviews (23)

View Details

Write a Review

Great Camera
 Written: Feb 06 11

Product Rating: ★★★★★

Pros: color of the camera
 the beautiful pictures it takes
 clearly
 weight of camera

Cons: nothing what-so-ever

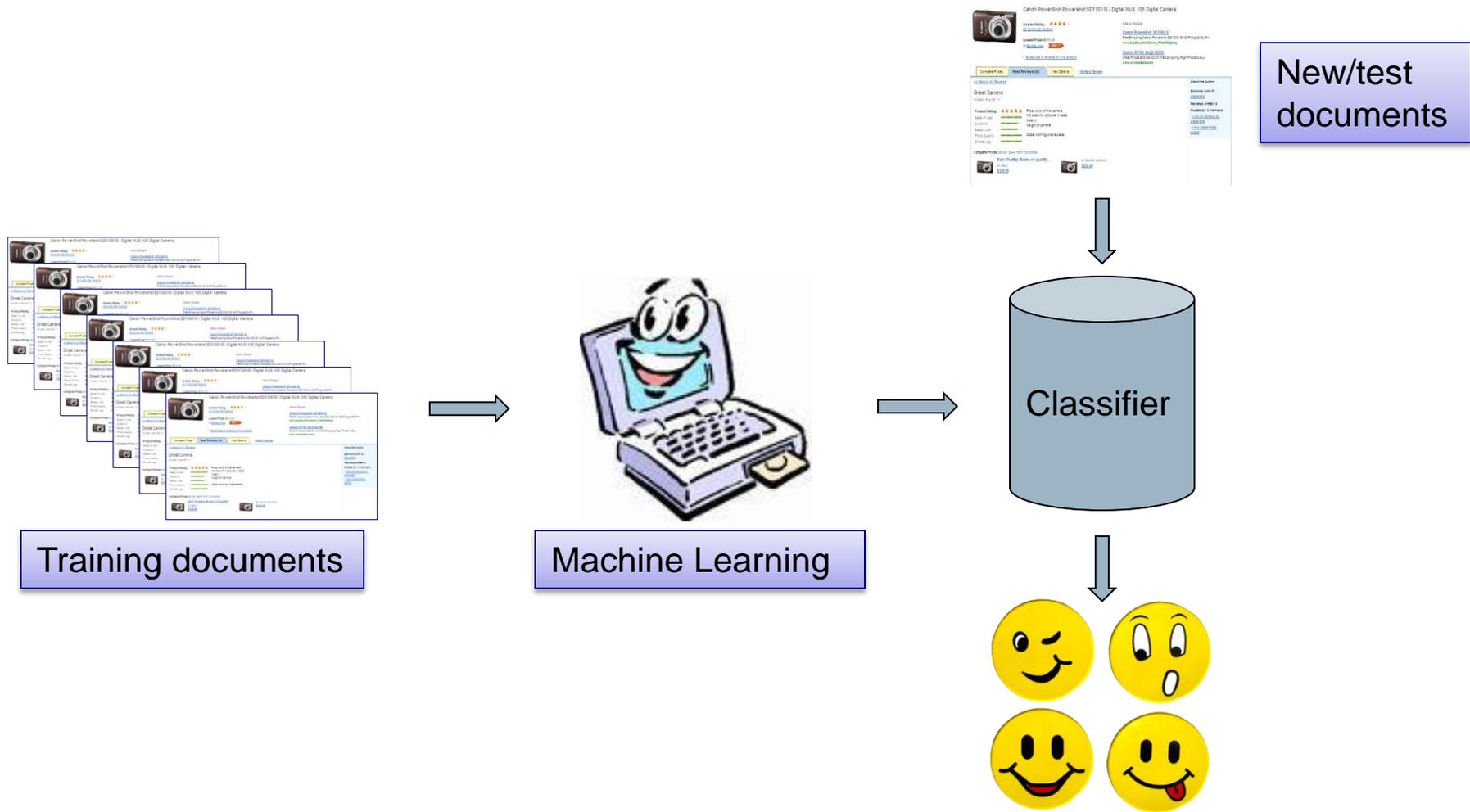
Compare Prices (\$109 - \$240 from 19 stores)

Earn 2% eBay Bucks on qualify...
 at eBay **\$129.99**

at Sports Authority **\$239.99**

About the Author
 Epinions.com ID:
[playerlanc](#)
 Reviews written: 9
 Trusted by: 0 members
[View all reviews by playerlanc](#)
[View playerlanc's profile](#)

Supervised Learning



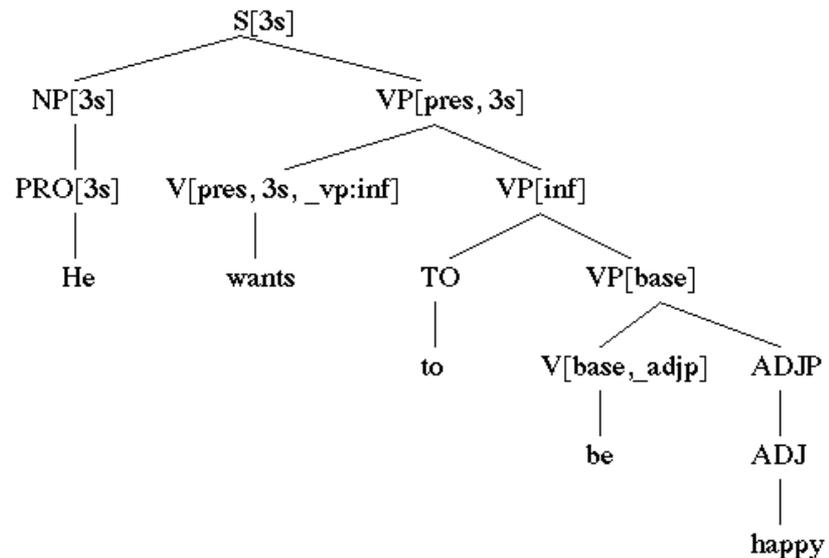
Supervised Learning (cont.)

- A *machine learning* technique: find patterns in known examples and apply to new documents
 - *Training* and *testing* examples
 - A set of data *features* to represent documents
- Learning goal: target *classes* (e.g. positive vs. negative)
- Product reviews domain is very common
 - Positive: 4-5 stars (thumbs up)
 - Negative: 1-2 stars (thumbs down)

Supervised Learning – Feature Extraction

- Terms and their frequency
 - Unigram and more general n-grams
 - Word position information
 - Term Frequency – Inverse Document Frequency (TFIDF) weighting
- Part of speech tags: adjectives are usually important indicators of subjectivities and opinions
- Syntactic dependency, e.g.
 - syntactic parsing tree

openNLP



- For a more comprehensive survey of attributes, see [Pang and Lee, FATIR2008]

Popular Supervised Learning Methods

- *Naïve Bayes* (NB):
 - A simple probabilistic classifier based on applying [Bayes' theorem](#) with strong (naive) independence assumptions
- *Maximum Entropy* (ME)
 - A probabilistic model that estimates the conditional distribution of the class label
- *Support Vector Machines* (SVM) [Pang et al, EMNLP2002]
 - A representation of the examples as points in space in which support vectors are computed to provide a best division of points/examples into categories
- *Logistic Regression* (LR) Model [Pang & Lee, ACL 2005]
 - A LR model predicts the classes from a set of variables that may be continuous, discrete, or a mixture

Domain Dependency Issue

- A classifier trained using opinionated documents from domain A often performs poorly when tested on documents from domain B
- Reason 1: words used in different domains can be substantially different, e.g.
 - *Cars* vs. *movies*;
 - *Cameras* vs. *Strollers*
- Reason 2: some words mean opposite in two domains, e.g.
 - “*unpredictable*” may be negative in a car review, but positive in a movie review [Turney, ACL2002]
 - “*cheap*” may be positive in a travel/lodging review, but negative in a toys review

Domain Adaptation

- **A well studied problem** [Aue & Gamon, RANLP2005; Blitzer et al, ACL 2007; Yang et al, TREC2006]
 - Step 1. Use labeled data from one domain and unlabeled data from both source the target domain and general opinion words as features
 - Step 2. Choose a set of pivot features which occur frequently in both domains
 - Step 3. Model correlations between the pivot features and all other features by training linear pivot predictors to predict occurrences of each pivot in the unlabeled data from both domains

Sentence-level Sentiment Analysis

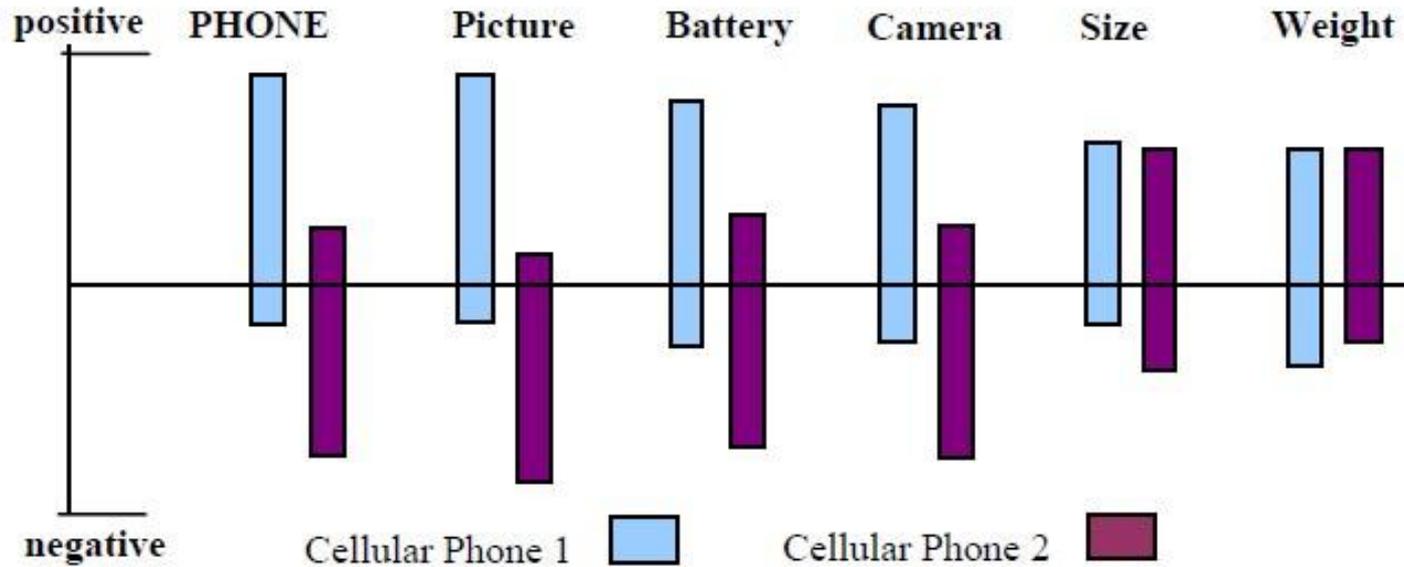
- Document level sentiment analysis is too coarse for most applications, 😊 or ☹️? e.g.
 - “I bought a new X phone yesterday. The voice quality is *super* and I really *like* it. However, it is a little bit *heavy*. Plus, the key pad is *too soft* and it *doesn't feel comfortable*. I think the image quality is *good* enough but I am *not* sure about the *battery life*...”
- Task: determine whether a sentence *s* is subjective or objective, and if *s* is subjective, determine whether its orientation is positive or negative
- Assumptions:
 - the sentence is opinionated on a single object
 - the opinion is from a single opinion holder

Learning Syntactic Patterns [Riloff & Wiebe, EMNLP2003]

- First, use high precision but low recall classifiers to automatically identify some subjective and objective sentences
 - A subjective classifier: the sentence contains two or more strong subjective clues
 - An objective classifier: the sentence contains no strong subjective clues
 - Based on manually collected single words and n-grams, which are good subjective clues
- Second, learn a set of patterns from subjective and objective sentences identified above
 - *Syntactic templates* are used to restrict the kinds of patterns to be discovered, e.g. <subject> active-verb → the customer complained
- Third, the learned patterns are used to extract more subjective and objective sentences (the process can be repeated)

Attribute-level Sentiment Analysis

- A positive/negative document does not mean the author likes/dislikes all attributes of the object [from Bing Liu's chapter on sentiment analysis]



- Attributes can be product properties, important topics, etc.

Topic Identification – Pre-defined

- Obtain topics from database, e.g. product properties (image size, quality, weight)
- Identify related topics, e.g.
 - “the pictures are very clear” → explicit attribute: **picture**
 - “It is small enough to fit easily in a coat pocket or purse” → implicit topic: **size**
- Group related topics, e.g.
 - Size <-> {large, small}
 - Weight <-> {heavy, light}
- Require domain knowledge (metadata)

Automated Topic Extraction – Examples

- *Term Frequency* (TF): favor frequently appearing words
- *Term Frequency – Inverse Document Frequency* (TFIDF): favor words that appear frequently in one document but relatively rarely in the corpus
- *Latent Dirichlet Allocation* (LDA): identify topics that are characterized by a distribution of words [Blei et al, JMLR2003]

How LDA Works

- Traditional model: document-term matrix (each document is a vector in the keyword space)

Terms	Documents													
	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14
abnormalities	0	0	0	0	0	0	0	1	0	1	0	0	0	0
age	1	0	0	0	0	0	0	0	0	0	0	1	0	0
behavior	0	0	0	0	1	1	0	0	0	0	0	0	0	0
blood	0	0	0	0	0	0	0	1	0	0	1	0	0	0
close	0	0	0	0	0	0	1	0	0	0	1	0	0	0
culture	1	1	0	0	0	0	0	1	1	0	0	0	0	0
depressed	1	0	1	1	1	0	0	0	0	0	0	0	0	0
discharge	1	1	0	0	0	1	0	0	0	0	0	0	0	0
disease	0	0	0	0	0	0	0	0	1	0	1	0	0	0
fast	0	0	0	0	0	0	0	0	0	1	0	1	1	1
generation	0	0	0	0	0	0	0	0	1	0	0	0	1	0
oestrogen	0	0	1	1	0	0	0	0	0	0	0	0	0	0
patients	1	1	0	1	0	0	0	1	0	0	0	0	0	0
pressure	0	0	0	0	0	0	0	0	0	1	0	0	0	1
rats	0	0	0	0	0	0	0	0	0	0	0	1	1	1
respect	0	0	0	0	0	0	0	1	0	0	0	1	0	0
rise	0	0	0	1	0	0	0	0	0	0	0	0	0	1
study	1	0	1	0	0	0	0	0	1	0	0	0	0	0

- New model:
 - A new set of vectors to span the keyword space, which can better represent each document
 - From matrix theory, we know singular vectors of documents are such ones, and the related singular values represent the **importance** of each singular vector
 - These singular vectors can be thought as unit vectors to define a new coordinate system
 - Keep the most important k singular vectors only
 - Dimensionality reduction: project the documents to the new k -d space.
- KEY: we combine the useful information from multiple keyword vectors and put it in one singular vector, so we need less vectors to represent the important information in the original collection

Examples of Extracted Topics

- A *topic* is characterized by a distribution over words

Ⓢ TOPIC_45	1154	0.129, read	0.075, hard	0.074, color	0.051, colors	0.050, blue	0.037, light	0.029, eyes
○ TOPIC_56	1150	0.160, category	0.123, categories	0.078, specific	0.043, general	0.034, sub	0.024, choose	0.022, example
○ TOPIC_53	1136	0.094, choices	0.075, window	0.061, top	0.056, button	0.040, annoying	0.040, pop	0.036, options
○ TOPIC_89	1115	0.594, search	0.106, advanced	0.058, feature	0.039, function			

TOPIC_45	0.673910	2883	feedback "WELL.....not so good. I found fire engine. I found fire truck. I did not find fire pumper. Nor fire apparatus. Thought that I'd leave some feedback, so I clicked on feedback. Guess what came up!!!! Fire Pumper. Three searches later, you get the results you're looking for. Still needs some work. Try again. Lettering is getting smaller, and finer line. It's hard to read. The light blue color looks very pale. If your color isn't set right, you see nothing. What's wrong with just BLACK print? Don't even try RED! Wonder what color blind persons think of the change. I think they see the light blue as next to non-existent. ___
TOPIC_45	0.595534	10227	colors "Light blue print is hard to read. That color, and most blues DO NOT Fall Into the NORMAL human color range for reading/type. This is why you will NOT see that color of type in over \$150-Billion dollars worth of magazine ad's every year..... Did you do any research about the brain, mind, color connection, human ease of use.....Hello?? Get it ? "
TOPIC_45	0.591019	20428	New Ebay Motors site "Price font size & color change; = GOOD. Blue ink item description == POOR! BLURRY TO READ. Poor page to font size ratio. Smaller font print size stretches auction item across page. Causes extra effort, & time, to read each auction. Lettering font size designed for a 17 yr. old with 20/10 vision to read. WHAT ARE YOU THINKING? "

Orientation Classification for Each Topic

- Determine orientation of the opinion on each topic in a sentence using any of the following:
 - Counting polarity words
 - Rule-based
 - Supervise learning
- Aggregate all the opinions on each topic

Comparative Sentiment Analysis

- Motivation: opinions on multiple objects
- Sentence-based analysis
- Comparatives, e.g.
 - “iPhone4 is better than Nexus”
 - “Movie-X is less emotional than Movie-Y”
 - “Bank-1's service is as good as Bank-2's service”
 - “Laptop-1 has a webcam, laptop-2 does not”
- Superlatives
 - “The battery life of this phone is the longest”
 - “This chair is the least stable”

Detection of Comparisons

- Comparative adjectives, adverbs
 - more, less, words ending in *-er*, etc
- Superlative adjectives, adverbs
 - most, least, words ending in *-est*, etc
- Keywords
 - same, similar, etc
- Learning patterns around keywords
 - specialized sequence classification algorithms used [Jindal & Liu, 2006]

Preferred Object Identification

- Objective of comparison
- Lexicon-based approach
 - lists of positive/negative words, comparatives
 - rules for combining them, e.g. decreasing comparative + negative word → positive opinion
 - domain knowledge, e.g. car-x gives higher mileage than car-y
- Areas of active research [Ganapathibhotla & Liu, 2008]

Outline

1. Introduction
2. Sentiment Identification & Classification
3. **Key Applications**
4. Examples
5. Conclusions

Key Sentiment Analysis Applications

- Opinion Summarization

- Opinion Search & Retrieval

twitrratr
TRACKING OPINIONS ON TWITTER

SEARCHED TERM: **@barackobama**

POSITIVE TWEETS: **213** NEUTRAL TWEETS: **887** NEGATIVE TWEETS: **39** TOTAL TWEETS: **1139**

18.70% POSITIVE

- my lovely wife gets to see @barackobama in indy for a 2nd time this morning. meanwhile once again i get stuck at work instead. life ain' ... (view)
- @barackobama (let al) not sure if you've read this, but good points to ponder, good policy to consider implementing: http://snurl.com/4mm8n (view)
- got a bunch of new followers since i followed @barackobama. now is a good time to mention i mostly plurk: http://www.plurk.com/user/anamacha (view)
- @barackobama barack, our next president. it's a great time to be alive!!! (view)
- just got a nice campaign donation for @barackobama http://my.barackobama.com/page/outreach/view/main/kimbafund (view)
- just watched a great (and very motivating) speech by @barackobama (view)
- @barackobama is in my hometown speaking at the arena in which my high school

77.88% NEUTRAL

- I live in PA, and am NOT voting for John McCain. Pass it on. @fakesarahpalm (view)
- @barackobama Hello Champ! The Presidency is the bag. I am on the phones for you in Poundridge, NY this Saturday. Respectfully yrs, Phoebe (view)
- @BarackObama Wondering What You Really Think of The Mainstream Media?? (view)
- @BarackObama what is the deal with Bart Simpson and The Family Guy donating to your campaign? (view)
- @BarackObama The BlogTV Insomniac Ironman challenge is now at 5 PM Eastern time. (view)
- I got better food phone banking for @masstranshow than I did for @barackobama. How odd is that???? (view)
- Here's hoping that my 1100th tweet will be a heartfelt congratulations to President @BarackObama and VP @Biden! (view)

3.42% NEGATIVE

- did my eyes deceive me or did i just see a @barackobama attack ad. i thought the messiah of washington didn't need negative ads. (view)
- ... and i ~~can't~~ be able to until late tonight because @barackobama and i are getting together to save the economy today. so stay hush okay. (view)
- just voted for @barackobama and damn it feels good! (view)
- you can't win so just quit @barackobama ...o wait wrong box click (view)
- oh no! @madow reporting that @barackobama has left the campaigns trail to be w/ his grandmother. :((view)
- @lowradloguy yeah, @barackobama came online and promised all his twitter followers seats in his administration. ~~too bad~~ you missed it. (view)
- @barackobama i can take several more weeks of senator moccain's attacks, but the american people cannot take 4 more years of ~~bad~~ policies. (view)

TOPICS BREAKING POPULAR SOURCES ALL TWITTER TIMEFRAME LAST 24H LAST 7D Start date

"hotel argonaut" AND "san francisco" +valence:positive -sentiment:profane

sentiment

- joy
- sadness
- acceptance
- disgust
- fear
- anger
- anticipation
- surprise
- positive
- negative
- profane
- unsafe

valence

- positive
- negative
- ambiguous
- general

tonality

negative positive

group by topics single list basic search

- Opinion Spam & Opinion Quality

Opinion Summarization

- Main objective: convert natural language text to structured summary to gain insights and consumer opinions
- Little academic research available but crucial to applications
 - E.g. movie review mining and summarization [Zhuang et al, CIKM2006]
- Visualization using bar chart or pie chart based on quintuples

Structured Opinion Summary

- An attribute-based summary of opinions on an object or multiple competing objects [Hu & Liu, KDD2004; Liu et al, WWW2005; Zhai et al, WSDM2011]

Cellular phone 1:

PHONE:

Positive: 125 <individual review sentences>

Negative: 7 <individual review sentences>

Feature: **voice quality**

Positive: 120 <individual review sentences>

Negative: 8 <individual review sentences>

Feature: **size**

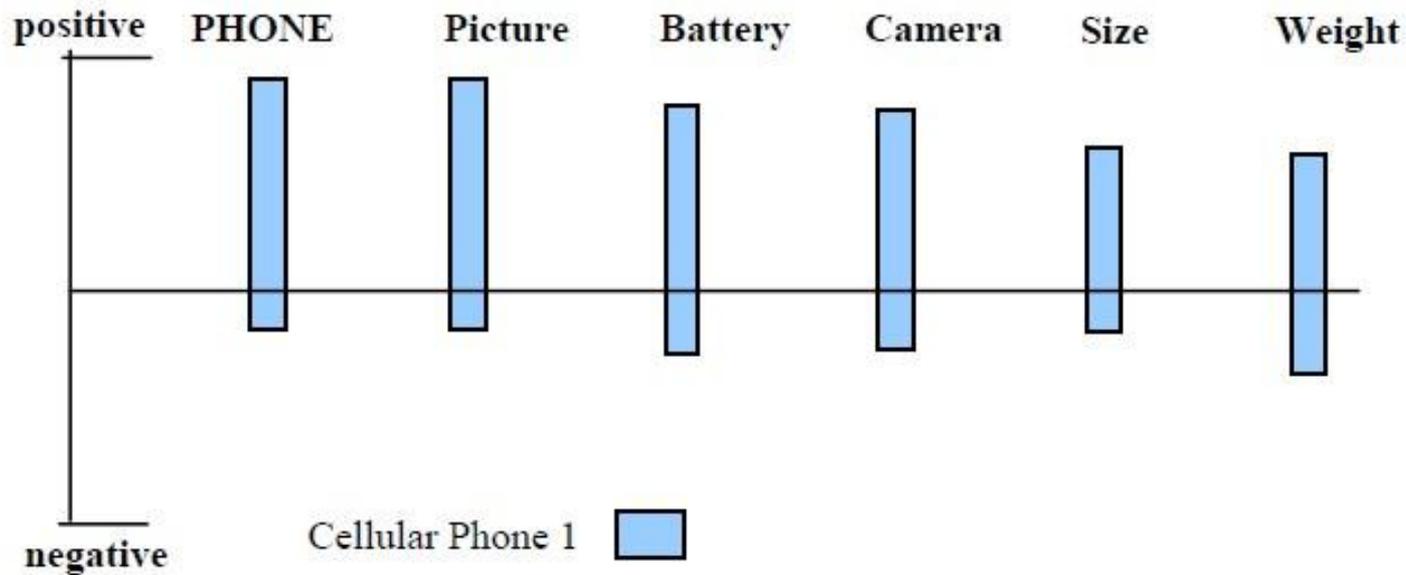
Positive: 80 <individual review sentences>

Negative: 12 <individual review sentences>

...

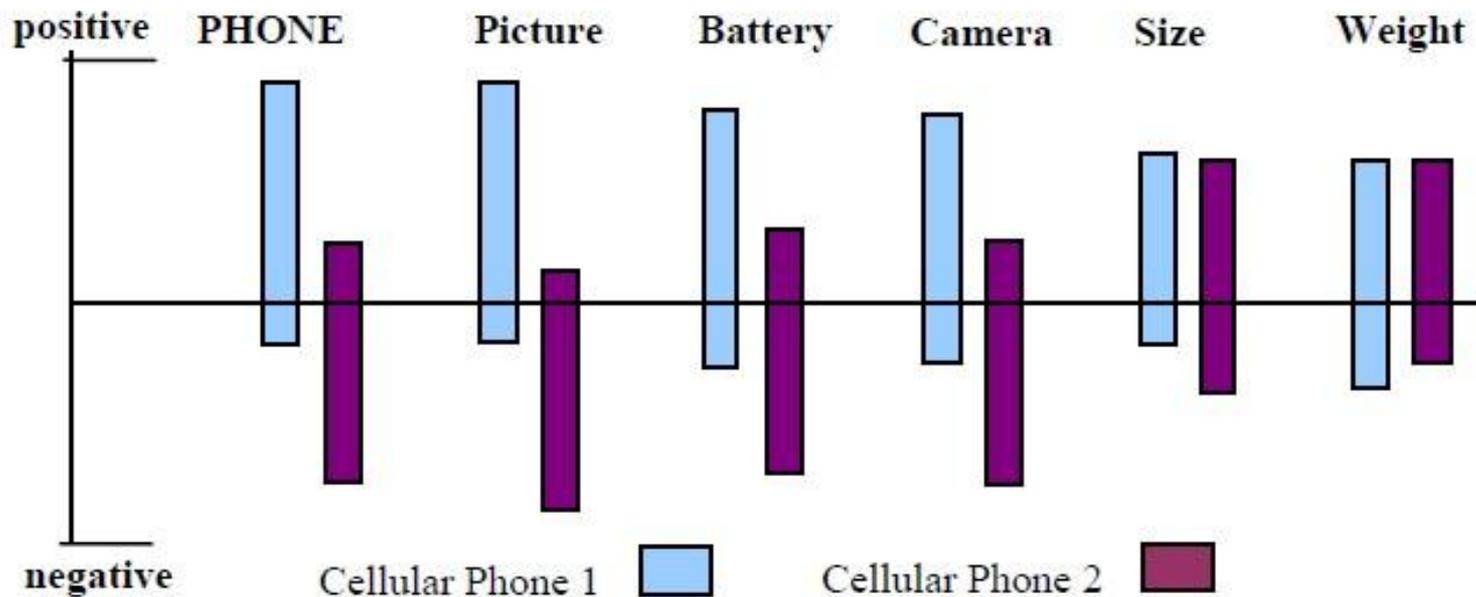
Visualization of an Attribute-based Summary

- More straightforward via visualization



Comparing Opinion Summaries of Multiple Products

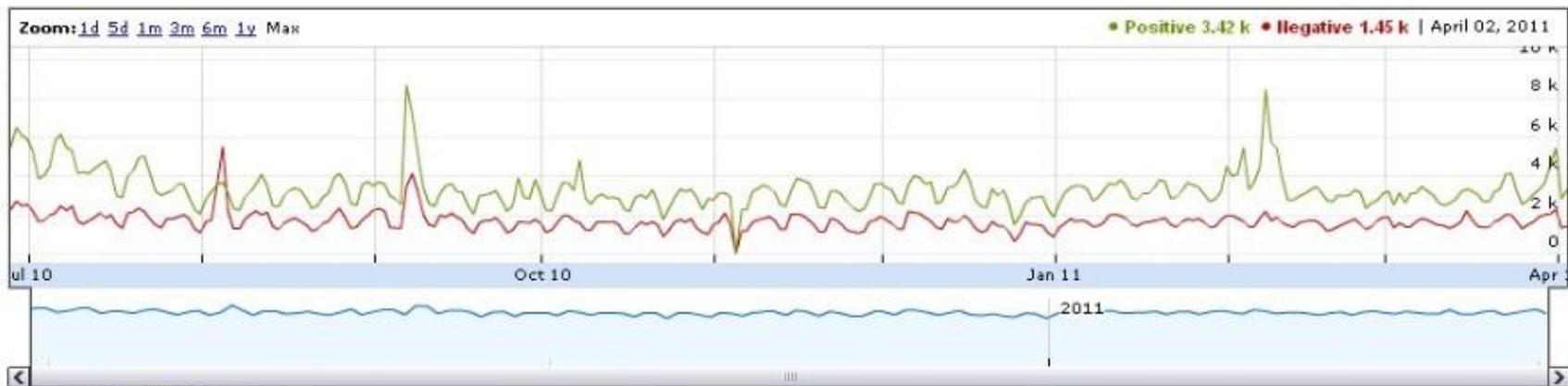
- More interesting to see a comparison between two or more competing products



- Many other visualization approaches are available [Pang & Lee, FATIR2008]

Summaries without Opinions

- Many types of summaries without opinions are also very useful and informative
- *Attribute buzz summary*: shows the relative frequency of attribute mentions, e.g. transaction security in an online banking study
- *Object buzz summary*: shows the frequency of mentions of different competing products
- *Sentiment trending*: reports the trend of sentiments over time



Sentiment Summarization Using Traditional Techniques

- Research in traditional text summarization is rich, see [Document Understanding Conference](#) (DUC)
- Produce a textual summary for a single document or multiple documents based on
 - *Abstraction*: generate natural language sentences using predefined templates, e.g. “Most people are positive about cell phone 1 and negative about cell phone 2”.
 - *Extraction*: keyword and key sentence extraction
- Limitation: qualitative but not quantitative

Opinion Search and Retrieval

- Find public opinions on a particular object or an attribute of the object, e.g. customer opinions on iPad2 or Japan nuclear crisis
- Find opinions of a person or organization (i.e. opinion holder) on a particular object or an attribute of the object, e.g. find Barack Obama's opinion on Japan nuclear crisis

Main Tasks

- *Data acquisition*: crawling and indexing Web documents
- *Document retrieval*: retrieving relevant documents/sentences to the query (same as conventional search engines)
- *Sentiment analysis*: determining whether the documents/sentences express opinions and whether the opinions are positive or negative (i.e. sentiment analysis)
- *Opinion ranking*: ranking opinionated documents according to certain criteria

Data Acquisition and Cleansing

- Web page crawling
 - Product reviews are usually easier to collect
 - Forum posts and blogs require more careful handling
- Web page parsing
 - Text extraction from Web pages
- Document indexing

Opinion Ranking

- Traditional Web search engines rank Web pages based on authority and relevance scores [Liu, WDM2006]
- Main objectives of opinion ranking:
 - Rank opinionated documents/sentences with high utilities or information contents at the top
 - Reflect the **natural distribution** of positive and negative opinions (same as in traditional opinion surveys)
- Simple solution: one ranking for positive opinions and one for negative ones, ratio of positive vs. negative as distribution
- **Attribute-based summary** for each opinion search is even better, unfortunately this is very difficult to achieve
- **Comparison search** will be useful too, e.g. opinions on Yahoo! mail vs. gmail vs. hotmail

An Opinion Ranking Example [Zhang & Yu, TREC2007]

- Document retrieval
 - In addition to keywords, certain concepts (named entities and various types of phrases, e.g. Wikipedia entries) are also used
 - Query expansion using synonyms of the search query and relevant words from top-ranked documents
 - Similarity/relevance score of each document with the expanded query is calculated using both keywords and concepts
- Opinion classification: SVM
 - Classifying each document into opinionated or not: subjective reviews from rateitall.com and epinions.com as well as objective information from Wikipedia are used as training data
 - Classifying each opinionated document into positive, negative, or mixed: reviews with star ratings from rateitall.com are used as training examples

Sentiment and Search

- http://www.nytimes.com/2010/11/28/business/28borker.html?_r=2&pagewanted=all shows how bad sentiments can improve search ranking!
- <http://googleblog.blogspot.com/2010/12/being-bad-to-your-customers-is-bad-for.html> gives Google's response!

A Bully Finds a Pulpit on the Web

By DAVID SEGAL

Published: November 26, 2010

SHOPPING online in late July, Clarabelle Rodriguez typed the name of her favorite eyeglass brand into [Google's](#) search bar.



David G. Klein

In moments, she found the perfect frames — made by a French company called Lafont — on a Web site that looked snazzy and stood at the top of the search results. Not the tippy-top, where the paid ads are found, but under those, on Google's version of the gold-medal podium, where the most relevant and popular site is displayed.

Ms. Rodriguez placed an order for both the Lafonts and a set of doctor-prescribed Ciba Vision contact lenses on that site, DecorMyEyes.com. The total cost was \$361.97.

It was the start of what Ms. Rodriguez would later describe as one of the most maddening and miserable experiences of her life.

RECOMMEND

TWITTER

COMMENTS (317)

SIGN IN TO E-MAIL

PRINT

REPRINTS

SHARE

WIN WIN
NOW PLAYING

Being bad to your customers is bad for business

12/01/2010 12:06:00 PM

A recent article by the New York Times related a [disturbing story](#). By treating your customers badly, one merchant told the paper, you can generate complaints and negative reviews that translate to more links to your site; which, in turn, make it more prominent in search engines. The main premise of the article was that being bad on the web can be good for business.

We were horrified to read about Ms. Rodriguez's dreadful experience. Even though our initial analysis pointed to this being an edge case and not a widespread problem in our search results, we immediately convened a team that looked carefully at the issue. That team developed an initial algorithmic solution, implemented it, and the solution is already live. I am here to tell you that being bad is, and hopefully will always be, bad for business in Google's search results.

As always, we learned a lot from this experience, and we wanted to share some of that with you. Consider the obvious responses we could have tried to fix the problem:

- Block the particular offender. That would be easy and might solve the immediate problem for that specific business, but it wouldn't solve the larger issue in a general way. Our first reaction in search quality is to look for ways to solve problems algorithmically.
- Use sentiment analysis to identify negative remarks and turn negative comments into negative votes. While this proposal initially sounds promising, it turns out to be based on a misconception. First off, the terrible merchant in the story wasn't really ranking because of links from customer complaint websites. In fact, many consumer community sites such as [Get Satisfaction](#) added a simple attribute called `rel=nofollow` to their links. The `rel=nofollow` attribute is a general mechanism that allows websites to tell search engines not to give weight to specific links, and it's perfect for the situation when you want to [link to a site without endorsing it](#). Ironically, some of the most reputable links to Decor My Eyes came from mainstream news websites such as the New York Times and Bloomberg. The Bloomberg article was about someone suing the company behind Decor My Eyes, but the language of the article was neutral, so sentiment analysis wouldn't have helped here either.

As it turns out, Google has a world-class sentiment analysis system ([Large-Scale Sentiment Analysis for News and Blogs](#)). But if we demoted web pages that have negative comments against them, you might not be able to find information about many elected officials, not to mention a lot of important but controversial concepts. So far we have not found an effective way to significantly improve search using sentiment analysis. Of course, we will continue trying.

- Yet another option is to expose user reviews and ratings for various merchants alongside their results. Though still on the table, this would not demote poor quality merchants in our results and could still lead users to their websites.

Related

[U.S. Arrests Online Seller Who Scared Customers](#) (December 7, 2010)

Add to Portfolio

- [Visa Inc](#)
- [Mastercard International Inc](#)

Opinion Search and Retrieval – Summary

- Being a TREC task since 2006, many approaches available [MacDonald et al, TREC2007]
- Main addition is the step of sentiment analysis
- Opinion ranking needs additional handling
- SVM classifiers have been popular among the best performers
- Research on comparison search is still limited

Opinion Spam

- Opinion spam: activities that try to deliberately mislead readers or automated opinion mining systems
 - *Hype spam*: giving undeserving positive opinions to some target objects in order to promote the objects, or
 - *Defaming spam*: giving unjust or false negative opinions to some other objects in order to damage their reputations
- Review spam is becoming a major issue, e.g. see <http://travel.nytimes.com/2006/02/07/business/07guides.html>
- Automatically detecting spam opinions becomes more and more critical

Different Types of Opinion Spam

- Bogus/fake opinions (hype, defaming reviews)
 - *“It's very expensive, however, I found www.yourfreeconsole.co.uk and got one for free! As you probably know, it's an expensive item for what it is... But, I didn't exactly pay for mine.”*
- Generic reviews: comment on brands, manufacturers or buyers/sellers, e.g.
 - *“Canon digital cameras are the best on earth...”*
- Non-opinions (advertisements, transactional text, random texts)
 - *“Get one for FREE!!! Have a look at this video first: http://www.youtube.com/watch?v=DFKYVE_Mug Just take 2 minutes and read this. This is very EASY to do!”*
 - *“Item was exactly the same as what was promised on sellers web page. Delivery was quick and professional. I am a very happy camper and would buy again in the future.”*

Opinion Spam Detection

- A binary classification problem [Jindal & Liu, WWW2007; Jindal & Liu, WSDM2008]
- For types 2 & 3, three sets of features are used
 - Review centric features: review text, #times that brands are mentioned, percentage of opinion words, review length, #helpful feedback, etc.
 - Reviewer centric features: average ratings given by a reviewer, standard deviation in rating, etc.
 - Product centric features: product price, sales rank, average rating, etc.
- Experimental results on product reviews seem promising

Opinion Spam Detection (cont.)

- For type 1, it is very difficult because
 - Manually labeling training data is very expensive
 - Spammers can craft a spam review that is just like any innocent review
- Research in this area is limited and in early stage, see [Jindal & Liu, WSDM2008] for more details

Opinion Quality Assessment

- Use consumer reviews of products as an example
- Usually posed as a regression task
 - Many review aggregation sites have buttons to collect user helpfulness feedback, which can be used for training and testing
 - Many types of data features are used [Kim et al, EMNLP2006; Ghose & Ipeirotis, ICEC2007; Zhang & Varadarajan, CIKM2006]
 - Review length
 - Review ratings (number of stars)
 - Counts of some specific POS tags
 - Opinion words and phrases
 - TF-IDF weighting scores
 - Product attribute mentions and product brands
 - Comparison with product specifications...
- Main application is in opinion search (ranking)

Useful Review vs. Spam

- Utility classification is different from spam detection
 - Not-helpful or low quality reviews are not necessarily fake reviews or spam
 - Helpful reviews may not be non-spam
- Users often determine the usefulness of a review based on whether it expresses opinions on many attributes of a product
 - Spammer can carefully craft a review that satisfies this requirement
 - *Feedback spam*: user feedbacks can be spammed too!
- A low quality review is still a valid review and should not be discarded, but a spam review is untruthful and/or malicious and should be removed once detected

Sentiment Analysis Is Everywhere!

Why sentiment analysis is the future of ad optimization

March 20, 2011 | Peter Yared

Comments 

 Like

 2 likes. Sign Up to see what your friends like.

 Tweet

[Peter Yared is the vice president of apps at Webtrends, which acquired Transpond, a social-apps developer he founded.]



Sentiment analysis is a hot new trend in social media, with the promise of helping brands understand what consumers are thinking and saying about their products. Products including early contender **Radian 6**, newcomers such as BuzzLogic, and my own company's Webtrends Social Measurement product are becoming pervasive in marketing organizations. But while consumer sentiment is important, what's much more important is revenue.

When revenue is down 10%, "but people like us!" is not an acceptable response from the head of marketing. Sentiment analysis isn't a solution unto itself, but it can be highly useful as a realtime feedback loop for advertising effectiveness and may soon be able to predict advertising results.

In the Mad Men-era heyday of mass marketing, marketing spend was impossible to quantify. TV, magazine, radio and billboard ads were purchased, and it was very difficult if not impossible to track exactly the return on investment of various slices of the marketing spend. Marketers would focus on issues like branding, messages and color schemes, with virtually no feedback loop other than the occasional focus group.

With the advent of digital marketing and online commerce, marketing spend and effectiveness is now tracked at the most minute level. How many ads are clicked on, how they convert, what is working and what is not is tracked at every level and segment. Even traditional legacy advertising is voraciously tracked, from Nielsen tracking how many consumers see a TV spot to **QR codes on billboards** acting as an effective realworld clickthrough.

In stark contrast to marketing of the past, today's marketers are measured by how much revenue they bring in per dollar spent. What "sentiment analysis" does is give those marketers an alternative way to measure their effectiveness — tracking how customers feel about and how much they are talking about a brand. All that effort on branding, messages, and color schemes can finally be validated!

<http://venturebeat.com/2011/03/20/why-sentiment-analysis-is-the-future-of-ad-optimization/> discusses ways to embed the analysis in an application!!

Roadmap

1. Introduction
2. Sentiment Identification & Classification
3. Key Applications
4. **Examples**
5. Conclusions

Examples – Two eBay Applications

- Application 1: Mining discussion forums
 - Vox Populi: Opinion Retrieval & Classification Engine
- Application 2: Enhancing eBay product reviews
 - Product Review Miner: Feature-based opinion mining
 - Product Reviews Spam Filtering

Mining eBay User Discussion Boards

Spring Seller Update

eBay teams responded to your questions about the **2011 Spring Seller Update** on a **special discussion board** on March 15th and 16th. This board is now available as a resource for more information around the announcements. Please read complete details about the changes on the **2011 Spring Seller Update Overview pages**.

Find and share information on your favorite categories -- from art to pottery & glass and much more. Search for keywords if you know exactly what you're looking for, or browse the discussion topics and see what you discover. Feel free to start your own topics too!

Boards	Views	Posts	Last Post
Animals	959,374	193,274	Apr 7, 2011 10:11 AM By desertdogscottie
Antiques	2,105,210	235,537	Apr 7, 2011 09:31 AM By lenorep
Art & Artist	479,202	101,629	Apr 7, 2011 09:18 AM By mario6489
Bears and Plush	56,468	16,924	Apr 6, 2011 10:09 AM By ammaafi
Book Readers	83,063	44,035	Apr 7, 2011 07:13 AM By scourgify
Booksellers	2,241,007	263,729	Apr 7, 2011 09:55 AM By indianfield

Product Related

Community & Technology Boards

Spring Seller Update

eBay teams responded to your questions about the **2011 Spring Seller Update** on a **special discussion board** on March 15th and 16th. This board is now available as a resource for more information around the announcements. Please read complete details about the changes on the **2011 Spring Seller Update Overview pages**.

Need help understanding an eBay policy? Have an eBay transaction you'd like advice on? eBay members answer questions, discuss best practices, and offer suggestions that help solve problems. Everyone has something to offer, so we encourage you to participate!

Boards	Views	Posts	Last Post
About Me page	2,414,562	15,785	Apr 7, 2011 05:52 AM By bernyhr007
Auction Listings	2,114,625	406,487	Apr 7, 2011 10:17 AM By suziblue52
Bidding	922,794	123,365	Apr 7, 2011 10:10 AM By 7606dennis
Buyer Central: Professional Buying	966,399	218,704	Apr 7, 2011 10:15 AM By snappyauctions14
Checkout	139,816	11,920	Apr 7, 2011 01:58 AM By skyerine21
eBay Mobile	50,719	996	Apr 5, 2011 02:17 PM By patrick240802
eBay Picture Hosting	72,643	3,860	Apr 2, 2011 09:35 AM By nates_tips
eBay Sidebar & Toolbar	32,864	1,055	Mar 29, 2011 01:30 AM By iamking100
eBay Stores	2,014,375	135,426	Apr 7, 2011 10:08 AM By motorgirl63

Technology Related

Vox Populi: Opinion Retrieval and Classification

Example: Health & Beauty Forum

- **Search:** Mineral Makeup
- ~ 700 posts – 202 opinions
- **Related Terms:** Mineral Makeup, bare mineral, estee lauder, bismuth oxychloride, bare essentials, everyday minerals, mineral veil, titanium dioxide
- **Search Results:** Positive, Negative, Neutral

Many of the newer mm companies are improving on the mineral makeup concept , and leaving out the bismuth

Mineral makeup is so daunting because what works for one might not work for another

Aromaleigh is the mineral makeup that I have liked best .

Everything I've read about Meow Mineral Makeup is positive so I'm really looking forward to trying them - I'll update as soon as I get them !

Example: Opinion Retrieval in Health & Beauty

Search: mineral makeup

Related Terms

The screenshot shows an eBay search results page for 'mineral makeup'. At the top, there is a search bar with 'mineral makeup' entered and a 'Search' button. To the right of the search bar, there are several related terms: 'mineral makeup', 'apply mineral', 'use bismuth', 'concealer brush', 'mineral companies', 'kabuki brush', 'baby kabuki', 'size brushes', and 'veena velvet'. Below the search bar, there is a list of search filters on the left, including 'what to put LOOSE MINERAL MAKEUP in?' (3/13) and 'Applying mineral makeup around eye area (?)' (9/29). The main content area shows a search result for 'Applying mineral makeup around eye area (?)' (9/29) by user 'gold*fish'. The post text discusses the use of bismuth in mineral makeup and mentions 'Monave' resellers. Below the post, there is a 'Topic Trend' section with two stacked bar charts. The left chart shows the distribution of 'Relevant' (orange) and 'Not Relevant' (green) posts across various topics. The right chart shows the distribution of 'Opinion' (orange) and 'No Opinion' (green) posts. A callout box highlights the 'Opinion Mineral Makeup ?' bar, showing 93 (25%) opinions out of a total of 367 posts.

Trends

Sentiment Classification: Experiment

- Data set
 - Human evaluation: 887 sentences - Two annotators
 - Annotations are consistent on 776 sentences (67 positive; 389 negative; 320 neutral)
- Vox Populi Opinion Classifier - 2008 Lab Results

	Positive	Negative	Neutral
Precision	59.6%	77.2%	78.2%
Recall	50.7%	86.1%	69.7%

Sarcasms in Technology discussion boards -> false positives

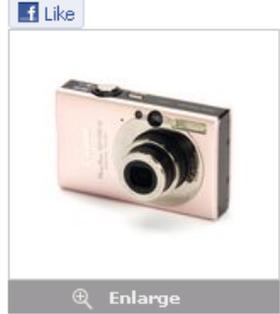
Product Review Miner

Feature-based Sentiment Analysis of Product Reviews

- Document level:
Unstructured product reviews: free text + User ratings (1 to 5 stars)
- Topic/Feature level
 - Ex: The **powershot SD1100** is **light** and **compact**
Camera *Weight* *Size*

Product Review Miner: Feature-based Opinion Mining

Canon PowerShot SD1100 IS Digital ELPH Pink Bundle 8.0 Megapixel Digital Camera



Manufacturer: Canon
Camera Type: Compact
Resolution (Megapixels): 8
Optical Zoom: 3
Product ID: EPID71198805

Description: Canon PowerShot SD1100 IS Digital ELPH Pink Bundle - Digital camera - compact - 8.0 Mpix - optical zoom: 3 x - supported memory: MMC, SD, SDHC, MMCplus - pink melody - with Canon SELPHY CP760 Compact Photo Printer [read more](#)

[Save this product](#) [Sell one](#)

[Write a review](#)

Portions of this page Copyright 1997 - 2011 CNET Channel. All rights reserved.

User Options

- General Comments**
 86% positive
- Ease Of Use**
 86% positive
- Photo Quality**
 89% positive
- Affordability**
 88% positive
- Features**
 96% positive
- Speed**
 83% positive

[Listings](#) [Product details](#) **48 reviews** ★★★★★

Buy this item

- Canon PowerShot SD1100 8.0 MP Digital Camera Package :)
\$76.99 (28 Bids)
 1h 57m
- *Canon PowerShot SD1100 8MP Digital Camera+WARANTY*
\$114.95
 7d 1h 28m
- Canon PowerShot SD1100 IS 8.0MP 3X Op Digital Camera
\$89.95
 26d 0h 47m

[See all 29 items](#)

Reviews [Write a review](#)

Sort by: Relevance: Most relevant first

Great little camera. Perfect for Digital Scrapbooking

★★★★★

Created: 11/09/09

by: [anditan83](#) (72 ★)

[See all reviews by anditan83](#)

I originally purchased a Canon Powershot SD1100 from Walmart a few years ago. I chose it only because it has the fastest shutter speed available. The more I used it, the more I loved it. One of my favorite features is COLOR ACCENT. It allows you to chose one color to keep and the rest of the picture is black and white. It has another feature that I found particularly useful is the COLOR SWAP. I was able to swap my white walls with various colors from paint chips from the stores and see INSTANTLY how it would look. Although I must note that while using these features you must keep the camera VERY still or they will blur. There is a little icon in the left hand corner that lets you know the camera senses minute movement. Using a mini tripod, resting your elbows on a flat surface, or just tucking them in close to your body will produce clear creative pictures.

All the pictures I took with this camera turned out crisp and clear, there was some graininess in low light situations without flash. My camera has survived a few falls, being used by my 8 year old, and four year old like I use

Product reviews



Structuring eBay Product Reviews (2)

Digital Camera Reviews
Categories ▾ Motors Stores **Deals**

EOS Rebel XSi 12.2 Megapixel > Search results



Canon PowerShot SD1100 IS Digital ELPH 8.0 Megapixel
[Sell one like this]

Manufacturer Part Number: 2756B001, 2757B001
Resolution (Megapixels): 12.2
Title: Canon PowerShot SD1100 IS Digital ELPH 8.0 Megapixel
Digital Camera Brand: Canon
Flash Type: Pop-up flash
ePID: 65880384

Description: For stunning photography with point-and-shoot ease, look no further than Canon's EOS Rebel XSi. The EOS Rebel XSi brings staggering technological innovation to the masses. It features Canon's EOS Integrated Cleaning System, Live View Fun... [read more](#)

Portions of this page Copyright 1997 - 2009 CNET Channel. All rights reserved.

Great!
SAVE 37%

Listings **Product details** **66 reviews**  4.5/5

User Options

- General Comments**  91% positive
- Ease Of Use**  87% positive
- Photo Quality**  89% positive
- Size**  97% positive
- Features**  96% positive
- Affordability**  90% positive

Buy this item

Canon EOS Rebel XSi 12.2 Megapixel
\$561.00 (10 Bids)
8h 31m

Canon EOS Rebel XSi Digital SLR Camera Body 5

Reviews

★★★★★ **great pocket camrea**
by: [eunine_u9](#) (50 ★)
8 of 10 people found this review helpful.

If you are looking for a portable , light weight , pocket , point and shoot camera , this is the one . It looks very nice and has a picture quality is excellent . I am very happy with the purchase !

Was this review helpful? [Report this review](#)

★★★★★ **Believe Me... you have the best in your hand!!**
by: [csu.ratul](#) (88 ★)
7 of 7 people found this review helpful.

Canon no doubt has proved time and again that it can give serious deal to Sony when it comes to Digital Camera . Looking months I decided to go for Canon sd 1100 and believe me it has made me wonder . " How can a camera be so beautiful ? " it has been giving me . The photos are beautiful . The face tracking is a funny feature . Everytime I go out I keep it handy to it has given me the same result everytime . Perfect !!!

Inspite of all this I would still give it a 4 upon 5 because it needs to do something about the interface front . Its cool but not tl more on the indoor photos .

More user options should be provided . The iso can be changed to auto or high .

Overall it has proven its brand name Canon and I am very happy about it .

Was this review helpful? [Report this review](#)

A Peek at Product Oriented Opinion Mining

- Vocabulary Acquisition

- Polarity keywords
- Domain/Product keywords

*The battery life and picture quality are great (+),
but the view finder is small (-) (* contextual orientation)*

- Lexical Patterns

- *The battery life is [not bad (-)] (+)*

<topic1><not><neg> -> (+)

- *This camcorder [would be great (+)](-) if the view finder was not so small*

<condition><pos><><if> → (-)*

Dictionary-based Vocabulary Acquisition

- Resources:

- The General Inquirer (<http://www.wjh.harvard.edu/~inquirer>)
- SentiWordNet (<http://patty.isti.cnr.it/~esuli/software/SentiWordNet>)

```
abiogenetic#a 0.0:0.125
abiotrophy#n 0.125:0.375
abjuration#n 0.0:0.75
abjure#v 0.0:0.75
ablate#v 0.0:0.125
ablaze#a 0.09375:0.125
able#a 0.03125:0.09375
abls#n 0.0:0.125
abnormal#a 0.125:0.4583333333333333
abnormality#n 0.15625:0.1875
abominable#a 0.125:0.5625
abominably#r 0.25:0.3125
abominate#v 0.0:0.125
abomination#n 0.041666666666666664:0.0833
abortifacient#a 0.0:0.5
```

← *Incomplete, does not capture*

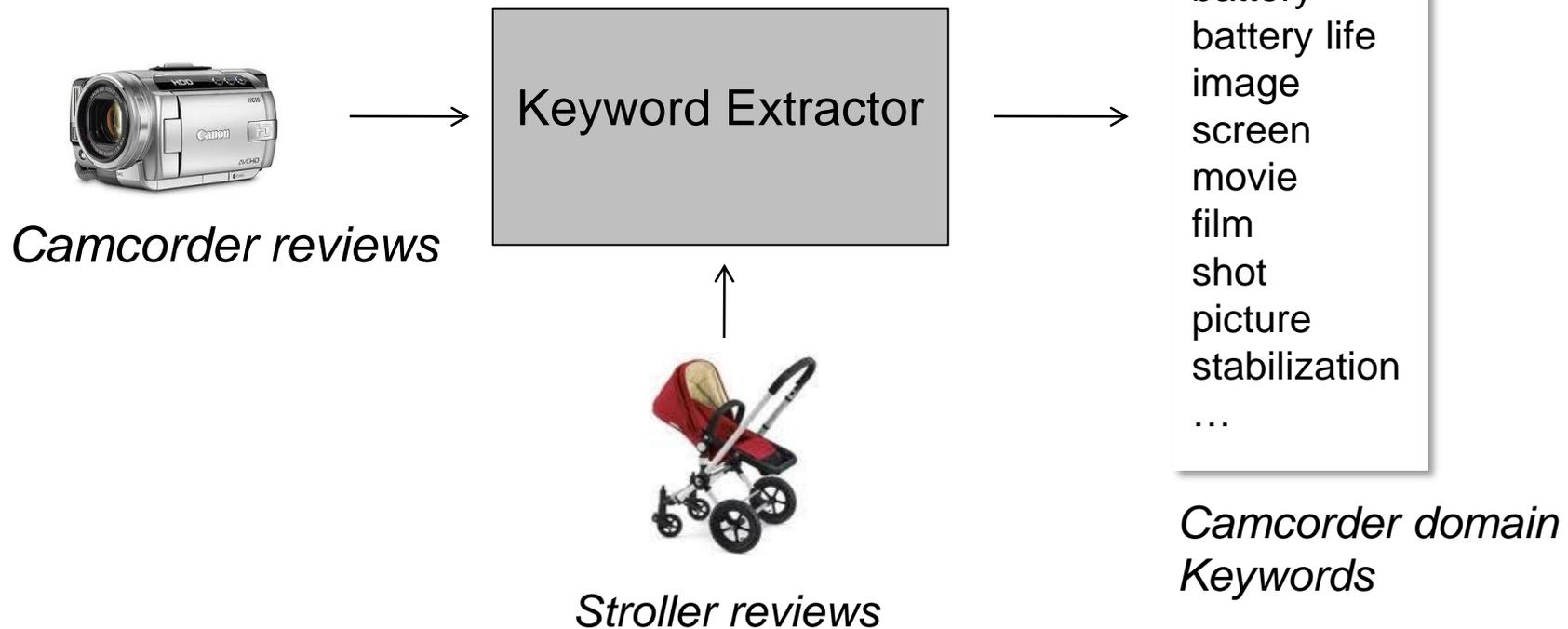
- *Domain vocabulary*
- *Emotional Expressions*
- *Typos*

Corpus-based Vocabulary Extraction

- Extracting Word Collocations:
 - *Examples: battery life, going back, turn off, etc.*
- Term Weighting: Extract domain and polarity terms
 - Learn from examples:
 - Product vocabulary: Compare different domains
 - Polarity words: Positive vs. Negative reviews

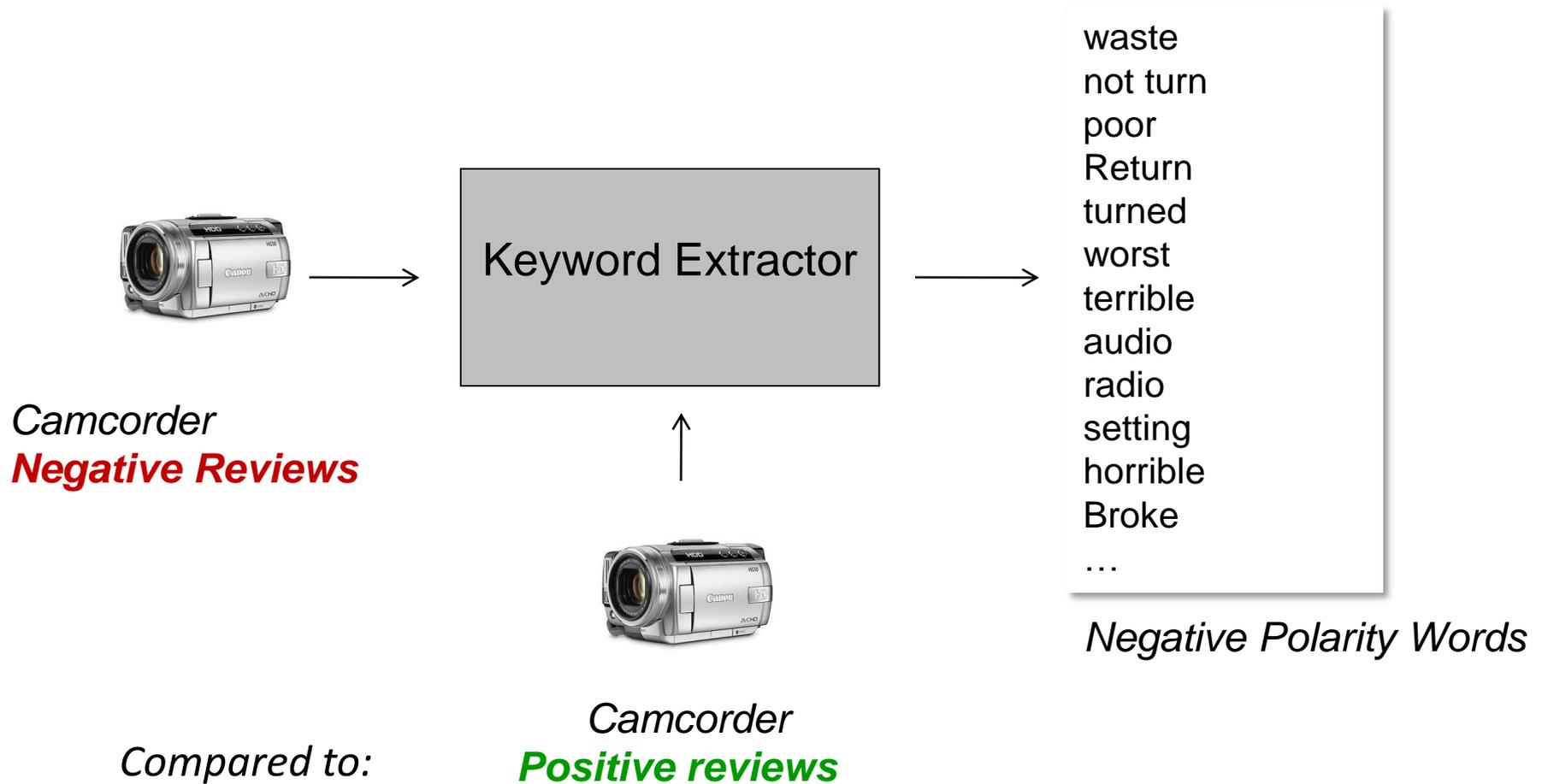
Vocabulary Extraction: Domain Vocabulary

Example: Camcorder Domain Vocabulary



Compared to "Not Camcorder"

Vocabulary Extraction: Polarity Keywords



Review Cleaning: Products vs. Sellers vs. Ads

Product Reviews

★★★★★

Item exactly what was promised on sellers webpage. Delivery was quick and professional. I am a very happy camper and would buy again in the future.

Seller Reviews

★★★★★

fast email responses and fast shipping!!

★★★★★

Horrible Customer Service

★★★★★

Great seller...Great service....

Spam

★★★★★

I found www.yourfreeconsole.co.uk and got one for free!

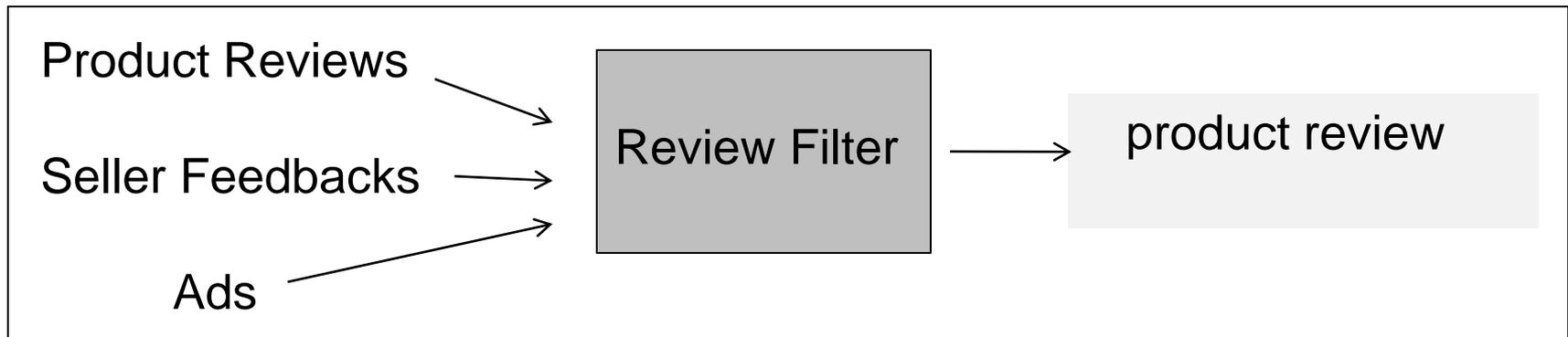
★★★★★

I went to www.gadgets.ipodgifts.net and signed up and complete one offer. I would recommend either The Total Credit Check or the eAuction tutor Offer they are both free and i kept the eAuction tutor offer to use with ebay. Another cool offer is the king.com.

get a FREE Playstation3! Not a SCAM! the real thing!

Product Review Quality Filtering

Goal: Overall product rating based only on product review ratings



Training

Classification

Automated Product Review Cleaning: Lab Tests (2008)

- Training: Manually annotated data sets

	Total Number	good	fair	bad	mixed	feedback	Spam
Books	536	58	184	167	46	80	1
DVDs	520	36	141	234	20	87	2
Video Game Systems	515	43	123	160	35	135	19

- Product Review Classification Results

			good + fair +bad → Review mixed + feedback → Seller Feedback
Cross Domain	Review	P	93.82
		R	97.56
	Seller Feedback	P	93.04
		R	82.60
	Ads	P	67.65
		R	54.76

(SVM Multi-Classification Model)

Roadmap

1. Introduction
2. Sentiment Identification & Classification
3. Key Applications
4. Examples
5. **Conclusions**

Sentiment Analysis Summary

Direct opinions (document-, sentence-, and attribute-levels)		Comparative opinions (sentence-level)
Counting polarity words	Use polarity dictionary (pre-defined or automatically learnt) to identify and classify sentiments	Use comparative and superlative words (mainly adjectives and adverbs) as well as special keywords (e.g. same, similar) to learn patterns around keywords. Specialized sequence classification algorithms are often used, e.g. [Jindal & Liu, 2006]
Rule-based	Use Label Sequential Pattern (LSP) Matching to define syntactic patterns for identification and classification of sentiments, e.g. [Chaovalit & Zhou, HICSS2005; Kim & Hovy, COLING2004; Turney, ACL2002; Yu & Hazivassiloglou, EMNLP2003]	
Supervised learning	Use machine learning techniques (Naïve Bayesian, Maximum Entropy, Support Vector Machines, and Logistic Regression) to build classifiers for identification and classification of sentiments, e.g. [Goldberg & Zhu, Workshop on TextGraphs, at HLT-NAAL2006; Mullen & Collier, EMNLP2004; Pang & Lee, ACL2005; Wiebe & Riloff, CICLing2005]	

Conclusions

- Sentiment Analysis tackles challenging tasks that involve NLP and text mining
- Being well studied yet many challenging problems not solved
- Strong commercial interest
 - Companies want to know how their products are being perceived
 - Prospective consumers want to know what existing users think
- Many start-ups emerging

Current Trends

- <http://www.opfine.com/> examines sentiment in market news
 - “buzz” is quantified and compared with other metrics such as price, trading-volume, etc.
 - Reliable and representative data sources are necessary

OPFINE
Opinionated financial news

Fri, 1 Apr 2011

Discovering sentiment in market news

OPFINE is statistical data mining engine of financial sentiment, markets sentiment and companies sentiment news from internet.
Opfine is fully automatic and has no human intervention.

Home Sign up Login About

Dow Jones sentiment

- Alcoa
- American Express
- Boeing
- Bank of America
- Citigroup
- Caterpillar
- Chevron
- Du Pont
- Walt Disney
- General Electric
- General Motors
- Home Depot
- Hewlett Packard
- IBM
- Intel
- Johnson & Johnson
- J.P. Morgan
- Kraft Foods
- Coca Cola
- Mc Donalds
- 3M
- Merck
- Microsoft

Opfine.com automatically analyses financial sentiment news from more than 3000 companies, 47 plus countries (and their relation to other countries), 20 plus commodities as well as detailed daily macro view, central banks and housing news. Most data with visual - graph based sentiment display. [Sign up](#) (simple, one step form to fill in) for a free account and see how it works in action.

Existing users can login [here](#)

Financial news sentiment from selected assets

365 days Macro news sentiment	365 days Commodities news sentiment	365 days Housing news sentiment
365 days generic Dollar news sentiment	365 days generic Euro news sentiment	365 days S&P 500 actual price

News sentiment graphs are based on 30 days sentiment aggregation.

Commodities popularity in the financial news in last 365 days

Commodity	Popularity
basemetals	74.0
wheat	174.0
platinum	62.0
preciousmetals	62.0
nickel	116.0
beef	117.0

Political Campaigns and Sentiments

- <http://www.nytimes.com/2010/11/01/technology/01sentiment.html> discusses applications in political campaigns
 - <http://www.feeltiptop.com/obama.php> examines sentiment trends to allow comparison with other timelines

Nation's Political Pulse, Taken Using Net Chatter

By JOSHUA BRUSTEIN
Published: October 31, 2010

When a Rasmussen poll last month showed Representative [Roy Blunt](#) opening a double-digit lead over Robin Carnahan in their Senate campaign in Missouri, John Hancock was not surprised.



Lorenzo Bevilacqua/CNN

John King of CNN. CNN will use sentiment analysis tools for the election on Tuesday.



Enlarge This Image

Mr. Hancock, a political consultant advising the Blunt campaign, had seen a similar shift in public opinion days earlier, through a software tool that analyzed the language being used in conversations about the campaign on social networking sites, blogs and other online conversations.

He said this technique, known as sentiment analysis, would soon be a part of every campaign he works on, because it helps him determine quickly which messages are resonating with potential voters. "You get a real sense of who's carrying the day," he said. "It affects the advice you're able to give."

TipTop SM
From:Life To:Life®

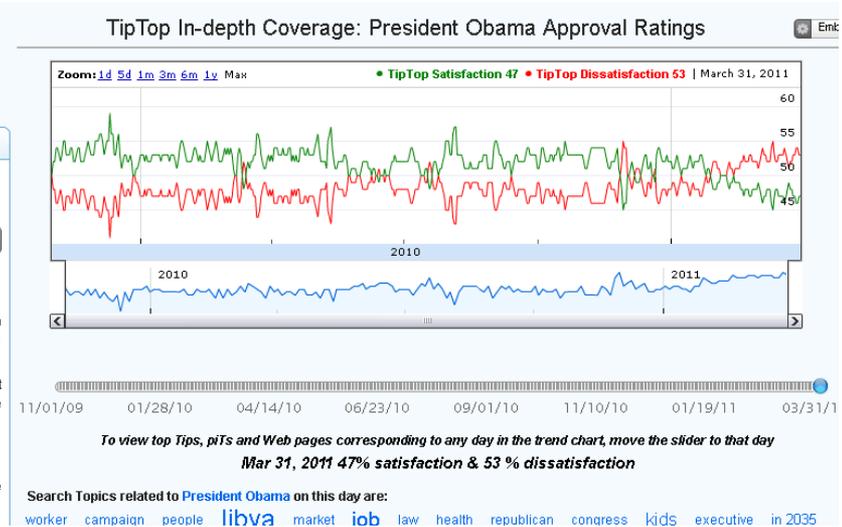
Search for Tips

Enter your search h

Search

Other TipTop In-depth coverage

- Visit [TipTop's Blog](#) to learn about the Obama Sentiment trend results
- What do TipTop Sentiment Trends, the Gallup Poll & the Dow Jones Index have in common? [Click here.](#)
- See how TipTop's satisfaction ratings compare to [Gallup's approval polls.](#)



Everyone Is Interested!!!

Exclusive: U.S. Spies Buy Stake in Firm That Monitors Blogs, Tweets

By Noah Shachtman  October 19, 2009 | 12:03 pm | Categories: [Info War](#), [Spies](#), [Secrecy and Surveillance](#)

America's spy agencies want to read your blog posts, keep track of your Twitter updates — even check out your book reviews on Amazon.

[In-Q-Tel](#), the investment arm of the CIA and the wider intelligence community, is putting cash into [Visible Technologies](#), a software firm that specializes in monitoring social media. It's part of a larger movement within the spy services to get better at using "open source intelligence" — information that's publicly available, but often hidden in the flood of TV shows, newspaper articles, blog posts, online videos and radio reports generated every day.



Visible crawls over half a million web 2.0 sites a day, scraping more than a million posts and conversations taking place on blogs, online forums, Flickr, YouTube, Twitter and Amazon. (It doesn't touch closed social networks, like Facebook, at the moment.) Customers get customized, real-time feeds of what's being said on these sites, based on a series of keywords.

- <http://www.wired.com/dangerroom/2009/10/exclusive-us-spies-buy-stake-in-twitter-blog-monitoring-firm/> reports on interest from intelligence agencies...

Future Research – Challenges...

- Identify implicit opinions, e.g.
 - *“I will never go back to any other camera”*
 - *“every person on Earth should own one of these”*
- Determine whether the subject and the object are relevant to the “true object”, e.g. in a phone review
 - “My mom was mad at me because I didn’t tell her I bought a new phone”
- Identify irony/sarcasm, e.g.
 - “Do you know there is a search engine called Google” (in a poor search context)
 - “This product is apparently designed by high school students”
- Standard data set for training/testing/evaluation is in great demand
- Social aspect: mine opinions from similar people only

Acknowledgements

- Bing Liu @ UIC
- Nitin Indurkha @ eBay Research Labs
- Sean Huang @ eBay Search Science
- Eric Brill @ eBay Research Labs
- Dennis Decoste @ eBay Research Labs

Online Resources

- Bing Liu. [Sentiment Analysis and Subjectivity](#), in *Handbook of Natural Language Processing, Second Edition* (editors: N. Indurkha and F. J. Damerau), 2010
- Bo Pang and Lillian Lee. [Opinion Mining and Sentiment Analysis](#), in *Foundations and Trends in Information Retrieval*, 2(1-2), pp. 1–135, 2008.
- Sentiment Analysis Symposium: <http://www.sentimentanalysisymposium.com/index.html>
- OpenNLP: <http://incubator.apache.org/opennlp/>
 - An organizational center for open source projects related to natural language processing, hosting a variety of java-based NLP tools which perform sentence detection, tokenization, pos-tagging, chunking and parsing, named-entity detection, and co-reference using the OpenNLP Maxent machine learning package.
- SentiWordNet: <http://sentiwordnet.isti.cnr.it/>
 - a polarity dictionary with strength scores.
- OpinionFinder in MPQA: <http://www.cs.pitt.edu/mpqa/>
 - a system that processes documents and automatically identifies subjective sentences as well as various aspects of subjectivity within sentences, including agents who are sources of opinion, direct subjective expressions and speech events, and sentiment expressions. OpinionFinder was developed by researchers at the University of Pittsburgh, Cornell University, and the University of Utah.
- LingPipe: <http://alias-i.com/lingpipe/>
 - a tool kit for text mining tasks such as sentiment analysis (<http://alias-i.com/lingpipe/demos/tutorial/sentiment/read-me.html>), named entity extraction, spelling correction, etc.
- VoxPop: <http://blog.typeslashcode.com/voxpath/>
 - A thesis project on sentiment analysis by [Andrew Mahon](#) and [Zeke Shore](#) at [Parsons School of Design](#), aiming to explore reader sentiment within comments of New York Times articles

References (1)

- A. Aue and M. Gamon. Customizing Sentiment Classifiers to New Domains: a Case Study. In Proceedings of Recent Advances in Natural Language Processing (RANLP), 2005.
- S. Baccianella, A. Esuli, and F. Sebastiani. SENTIWORDNET 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC), pp. 2200-2204, 2010.
- D. Blei, A. Ng, and M. Jordan. Latent Dirichlet Allocation. *Journal of Machine Learning Research (JMLR)*, 3, pp. 993-1022, 2003.
- J. Blitzer, M. Dredze, and F. Pereira. Biographies, Bollywood, Boom-boxes and Blenders: Domain Adaptation for Sentiment Classification. In Proceedings of the Association for Computational Linguistics (ACL), 2007.
- Eric Brill. A Simple Rule-Based Part Of Speech Tagger. In Proceedings of the Third Conference on Applied Natural Language Processing (CANLP), pp. 152-155, 1992.
- P. Chaovalit and L. Zhou. Movie Review Mining: a Comparison between Supervised and Unsupervised Classification Approaches. In Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS'05) , 2005.
- K. Dave, S. Lawrence, and D. M. Pennock. Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews,” *Proceedings of WWW*, pp. 519–528, 2003.
- C. Fellbaum, eds. *Wordnet: An Electronic Lexical Database*. MIT Press, 1998.
- M. Gamon. Sentiment Classification on Customer Feedback Data: Noisy Data, Large Feature Vectors, and the Role of Linguistic Analysis,” *Proceedings of the International Conference on Computational Linguistics (COLING)*, 2004.
- M. Gamon, A. Aue, S. Corston-Oliver, and E. Ringger. Pulse: Mining Customer Opinions from Free Text. In Proceedings of the International Symposium on Intelligent Data Analysis (IDA), pp. 121–132, 2005.
- A. Ghose and P. G. Ipeirotis. Designing Novel Review Ranking Systems: Predicting Usefulness and Impact of Reviews. In Proceedings of the International Conference on Electronic Commerce (ICEC), 2007.

References (2)

- V. Hatzivassiloglou and K. McKeown. Predicting the Semantic Orientation of Adjectives. In Proceedings of the Joint ACL/EACL Conference, pp. 174–181, 1997.
- V. Hatzivassiloglou and J. Wiebe. Effects of Adjective Orientation and Gradability on Sentence Subjectivity. In Proceedings of the International Conference on Computational Linguistics (COLING), 2000.
- M. Hu and B. Liu. Mining and Summarizing Customer Reviews. In Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), pp. 168–177, 2004.
- N. Jindal and B. Liu. Review Spam Detection. In Proceedings of WWW, 2007.
- N. Jindal and B. Liu. Opinion Spam and Analysis. In Proceedings of the Conference on Web Search and Web Data Mining (WSDM), pp. 219–230, 2008.
- S. Kim and E. Hovy. Determining the Sentiment of Opinions. In Proceedings of the International Conference on Computational Linguistics (COLING), 2004.
- S. Kim, P. Pantel, T. Chklovski, and M. Pennacchiotti. Automatically Assessing Review Helpfulness. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 423–430, Sydney, Australia, July 2006.
- B. Liu. Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data. Springer, 2006.
- B. Liu, M. Hu, and J. Cheng. Opinion Observer: Analyzing and Comparing Opinions on the Web. In Proceedings of WWW, 2005.
- Y. Lu, M. Castellanos, U. Dayal, and C. Zhai. Automatic construction of a context-aware sentiment lexicon: An optimization approach. In Proceedings of 20th International World Wide Web Conference, 2011.
- T. Mullen and N. Collier. Sentiment Analysis Using Support Vector Machines with Diverse Information Sources. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2004.
- V. Ng, S. Dasgupta, and S. Arifin. Examining the Role of Linguistic Knowledge Sources in the Automatic Identification and Classification of Reviews,” Proceedings of the COLING/ACL Main Conference Poster Sessions, pp. 611–618, Sydney, Australia, July 2006.

References (3)

- B. Pang and L. Lee. Seeing Stars: Exploiting Class Relationships for Sentiment Categorization with Respect to Rating Scales. In Proceedings of the Association for Computational Linguistics (ACL), pp. 115–124, 2005.
- B. Pang and L. Lee. Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval 2(1-2), pp. 1–135, 2008.
- B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up? Sentiment Classification Using Machine Learning Techniques. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 79–86, 2002.
- P. Turney. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In Proceedings of the Association for Computational Linguistics (ACL), pp. 417–424, 2002.
- J. Wiebe, R. F. Bruce, and T. P. O'Hara. Development and Use of a Gold Standard Data Set for Subjectivity Classifications. In Proceedings of the Association for Computational Linguistics (ACL), pp. 246–253, 1999.
- T. Wilson, J. Wiebe, and R. Hwa. Just How Mad Are You? Finding Strong and Weak Opinion Clauses,” Proceedings of AAAI, pp. 761–769, 2004.
- T. Wilson, J. Wiebe, and P. Hoffmann. Recognizing Contextual Polarity in Phrase-level Sentiment Analysis. In Proceedings of the Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP), pp. 347–354, 2005.
- H. Yang, L. Si, and J. Callan. Knowledge Transfer and Opinion Detection in the TREC2006 Blog Track. In Proceedings of TREC, 2006.
- H. Yu and V. Hatzivassiloglou. Towards Answering Opinion Questions: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences,” Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2003.
- Z. Zhai, Bing Liu, Hua Xu, and Peifa Jia. Clustering product features for opinion mining. In Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, 2011.
- Z. Zhang and B. Varadarajan. Utility Scoring of Product Reviews. In Proceedings of the ACM Conference on Information and Knowledge Management (CIKM), pp. 51–57, 2006.
- L. Zhuang, F. Jing, and X. Zhu. Movie review mining and summarization. In Proceedings of the 15th ACM Conference on Information and Knowledge Management (CIKM), 2006.