

Sentiment Analysis

MIE223
Winter 2025

1 Sentiment Analysis

1.1 Positive or negative movie review?

- unbelievably disappointing
- Full of zany characters and richly applied satire, and some great plot twists
- this is the greatest screwball comedy ever filmed
- It was pathetic. The worst part about it was the boxing scenes.

1.2 Google Product Search



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

Reviews

Summary - Based on 377 reviews



What people are saying

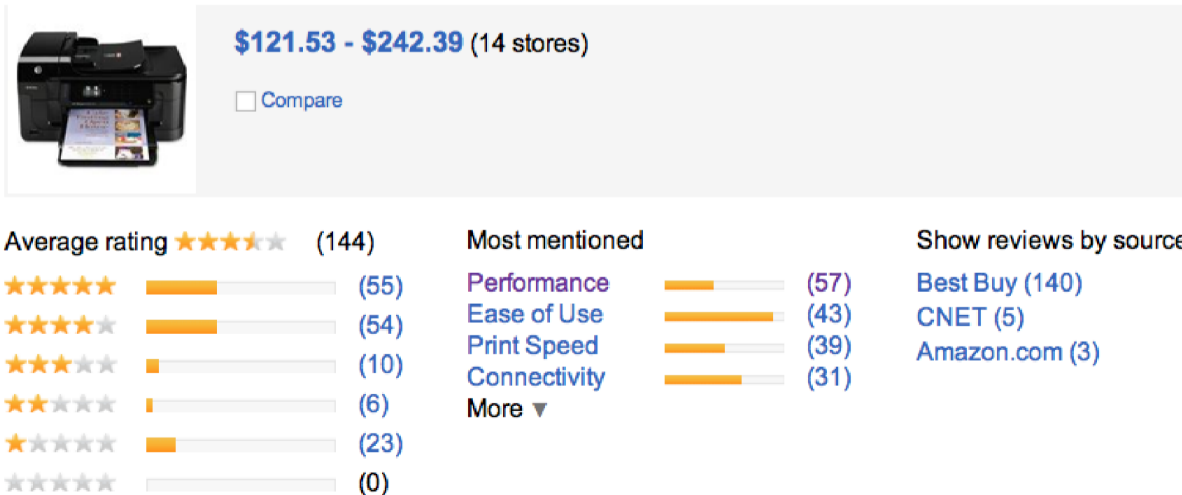
ease of use	<div><div></div></div>	"This was very easy to setup to four computers."
value	<div><div></div></div>	"Appreciate good quality at a fair price."
setup	<div><div></div></div>	"Overall pretty easy setup."
customer service	<div><div></div></div>	"I DO like honest tech support people."
size	<div>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

How do we find the sentiment of a review? you can automatically extract common tokens and phrases from the reviews and use them to determine the sentiment of the review. We use phrase chunking to get key phrases. use entailment to determine if a sentence is relevant to a specific key phrase. determine whether phrase is positive or negative called polarity analysis.

1.3 Bing Shopping

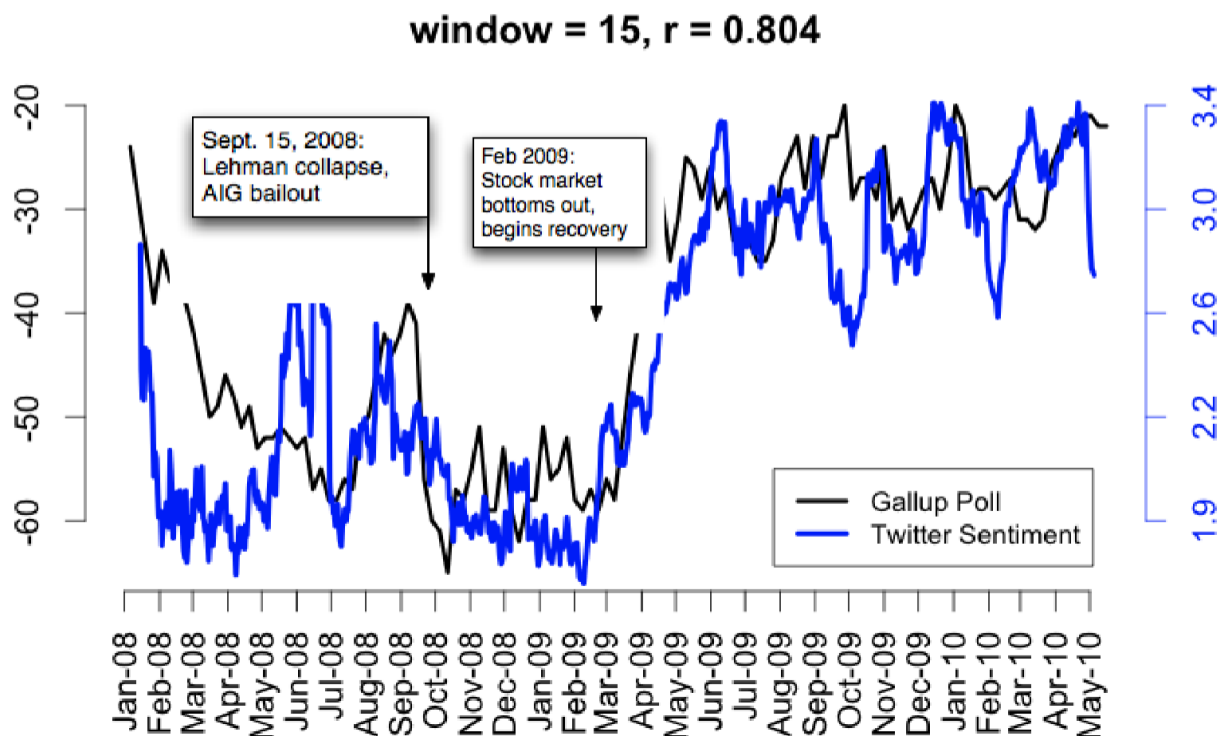
HP Officejet 6500A E710N Multifunction Printer

[Product summary](#) [Find best price](#) [Customer reviews](#) [Specifications](#) [Related items](#)



sentiment analysis is heavily used for product reviews

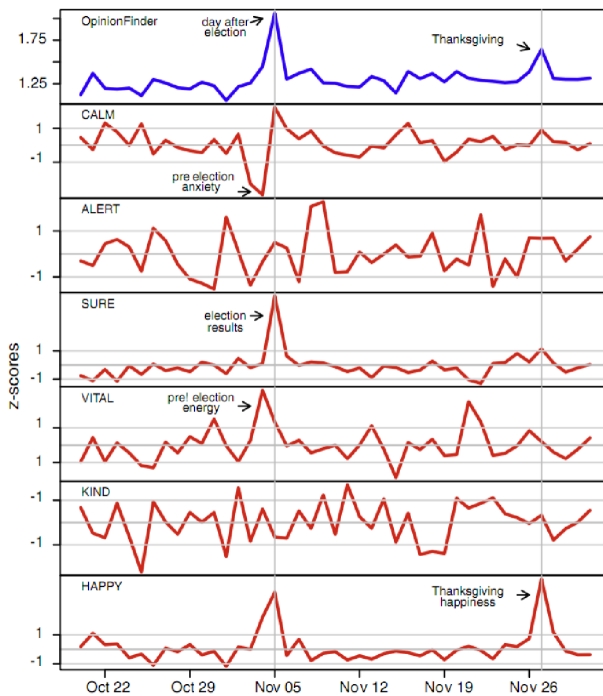
1.4 Twitter sentiment versus Gallup Poll of Consumer Confidence



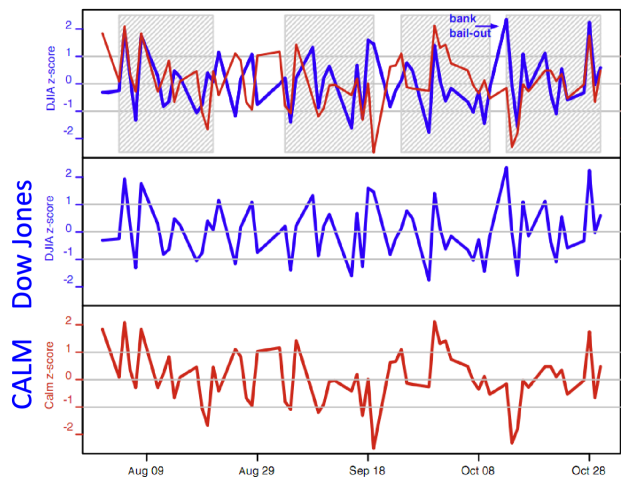
- gallup poll can be used to survey public opinion

- twitter sentiment can be used to predict public opinion

1.5 Twitter sentiment:

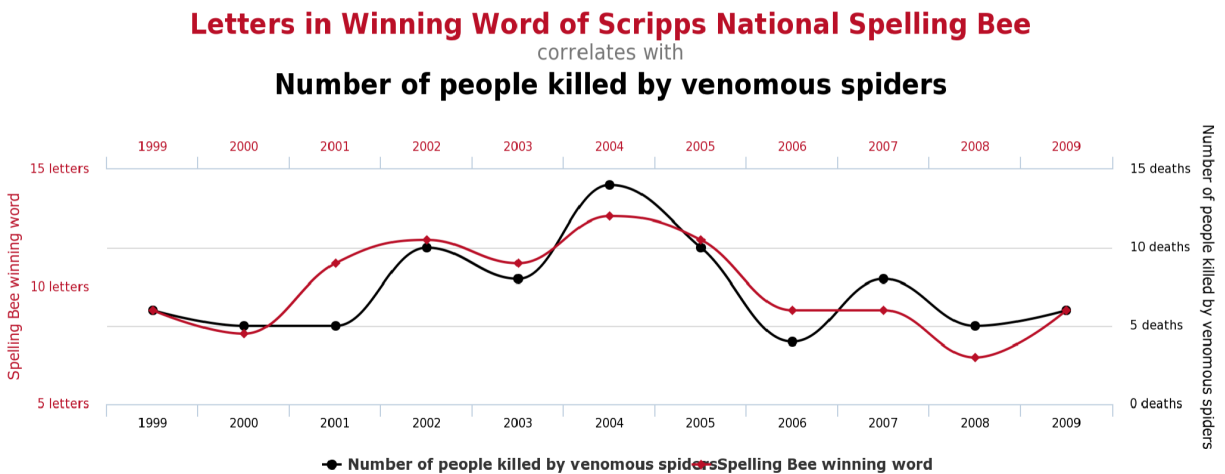


- CALM predicts DJIA 3 days later
- At least one current hedge fund uses this algorithm



- an efficient market makes use of information
- twitter sentiment can be used to predict stock prices
- the results are terrible
- ideas are good, nuanced emotions can influence the stock market
- different events give rise to different emotions which can be measured

1.6 Beware: Spurious Correlations



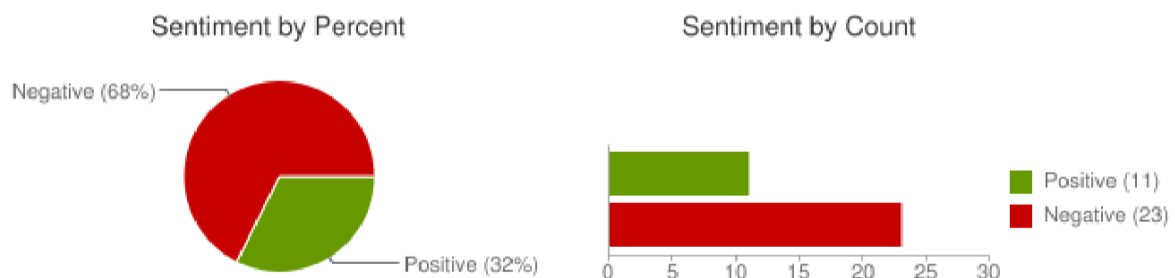
- data matching correlations are similar but not true

1.7 Target Sentiment on Twitter

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

[Save this search](#)

Sentiment analysis for "united airlines"



jlljacobson: OMG... Could @United airlines have worse customer service? W8g now 15 minut
[Posted 2 hours ago](#)

12345clumsy6789: I hate United Airlines Ceiling!!! Fukn impossible to get my conduit in this d
[Posted 2 hours ago](#)

EMLandPRGbelgiu: EML/PRG fly with Q8 united airlines and 24seven to an exotic destination
[Posted 2 hours ago](#)

CountAdam: FANTASTIC customer service from United Airlines at XNA today. Is tweet more
[Posted 4 hours ago](#)

1.8 Sentiment analysis has many other names

- Opinion extraction
- Opinion mining
- Sentiment mining
- Subjectivity analysis

1.9 Why sentiment analysis?

- Movie: is this review positive or negative?
- Products: what do people think about the new iPhone?
- Public sentiment: how is consumer confidence? Is despair increasing?
- Politics: what do people think about this candidate or issue?
- Prediction: predict election outcomes or market trends from sentiment

1.10 Scherer Typology of Affective States

- Emotion: brief organically synchronized ... evaluation of a major event
- angry, sad, joyful, fearful, ashamed, proud, elated
- Mood: diffuse non-caused low-intensity long-duration change in subjective feeling
- cheerful, gloomy, irritable, listless, depressed, buoyant
- Interpersonal stances: affective stance toward another person in a specific interaction
- friendly, flirtatious, distant, cold, warm, supportive, contemptuous
- Attitudes: enduring, affectively colored beliefs, dispositions towards objects or persons
- liking, loving, hating, valuing, desiring
- Personality traits: stable personality dispositions and typical behavior tendencies
- nervous, anxious, reckless, morose, hostile, jealous

Note 1. Majority of review sentiment focuses on attitudes. Public mood predicts stock market

1.11 Sentiment Analysis

- Sentiment analysis is the detection of attitudes
- “enduring, affectively colored beliefs, dispositions towards objects or persons”

1. Holder (source) of attitude
2. Target (aspect) of attitude

3. Type of attitude

- From a set of types like, love, hate, value, desire, etc.
- Or (more commonly) simple weighted polarity:
 - positive, negative, neutral, together with strength

4. Text containing the attitude like sentence or entire document

- Simplest task:
 - Is the attitude of this text positive or negative?
- More complex:
 - Rank the attitude of this text from 1 to 5
- Advanced:
 - Detect the target, source, or complex attitude types

2 Text Processing and Sentiment Classifier Limitations

2.1 Sentiment Classification in Movie Reviews

- Polarity detection:
 - Is an IMDB movie review positive or negative?

2.2 IMDB data in the Pang and Lee database

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

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

cool .

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

19



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

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

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

2.3 Sentiment Analysis and Prediction

- Sentiment-aware tokenization
- Important if text analysis should be sensitive to sentiment
- Prediction using different classifiers / regressors
- Could learn with Naïve Bayes, Logistic Regression, SVM
 - Vader is an off-the-shelf tool for predicting sentiment
 - Actually a regressor, predicts strength of polarity [-1,+1]
 - +1 = Most positive, -1 = Most negative, 0 = Neutral

2.4 Sentiment Tokenization Issues

Let's consider tokenization text analysis should be sensitive to sentiment

- Capitalization
 - Preserve capital letters
 - Why?
- Emoticons
- Useful code Regex

Christopher Potts emoticons

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

2.5 Classification Results

- Trained classifier achieves 92.1% accuracy on movies
- Off-the-shelf Vader would do well: around 80-85% accuracy
- Need to train a classifier per domain for best results

- Scary movie = good!
- Scary hotel = bad!
- Hotel with “thin walls”?

2.6 Classifiers don’t capture everything

Subtlety:

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