



Sentiment Analysis

Andrea Esuli



What is Sentiment Analysis?

What is Sentiment Analysis?

“Sentiment analysis and opinion mining is the field of study that analyzes people’s opinions, sentiments, evaluations, attitudes, and emotions from written language.”

[Bing Liu, “Sentiment Analysis and Opinion Mining” Morgan & Claypool Publishers, 2012.](#)

SA works on the **subjective/evaluative/emotive** components of textual information, which have often been ignored in the **objective/factual/topical** analysis usually performed in traditional TA.

Topic vs Sentiment

Topic and sentiment are two main orthogonal dimensions:

- Topic/Fact/Objective information
- Sentiment/Opinion/Subjective information (affective states, emotions. . .)

Topical analysis:

- Discriminating political news from sport news.
- Extracting mention of names of persons in text.

Sentiment analysis:

- Discriminating between favorable and negative attitude toward a subject.
- Identifying the expressions of an emotion and the target of that emotion.

Topic vs Sentiment

Objective information:

*The **4.7-inch** display on the iPhone 6 is arguably its best feature.*

*...concerns have been raised about the relatively low resolution (**1334 x 750 pixels**)*

[Source](#)

Topic vs Sentiment

Subjective information:

*The 4.7-inch display on the iPhone 6 is arguably its **best feature**.*

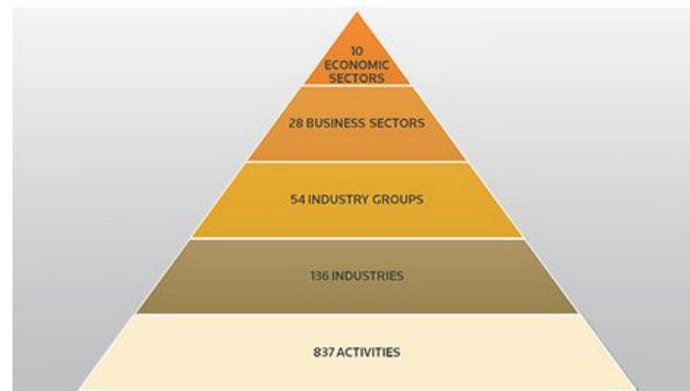
*...concerns have been raised about the **relatively low** resolution (1334 x 750 pixels)*

[Source](#)

Topic vs Sentiment

Classification of documents:

- with respect to the Thomson Reuters taxonomy*.



- with respect to the content being a positive, neutral, or a negative evaluation†.

```
{"data": [{"text": "I love Titanic.", "id":1234, "polarity": 4},  
          {"text": "I hate Titanic.", "id":4567, "polarity": 0}]}
```

* [Source](#) † [Source](#)

Topic vs Sentiment

Extraction of information:

- regarding objective properties

The **NBA** player **Michael Jordan** is from the **United States of America***

Organization **Person** **Location**

- regarding the expression of opinions.

soldiers with 20 years or more service are **generally satisfied** with **termination packages being offered**†

Agent **Attitude** **Target**

* [Source](#) † [Source](#)

L'esame È stato eseguito con sequenza T2 STIR e con sequenze T1 3D dinamiche prima e dopo somministrazione di mdc paramagnetico, acquisite secondo piani di scansione assiali.

Diffusi esiti cicatriziali in sede retroareolare sinistra.

Non si apprezzano potenziamenti sospetti bilateralmente.

In particolare non si È osservato un corrispettivo RM dell'immagine descritta mammograficamente a sinistra.

In sede ascellare bilateralmente si apprezzano alcune linfadenopatie di verosimile significato reattivo.

In relazione al quadro RM, si ritiene sufficiente eseguire controllo con esame ecografico tra 6 mesi.

CODICE ACR:06.1

Esiti chirurgici

BIRADS

Enhancement descrizione

Enhancement presenza/assenza

Indicazioni Esame

Informazioni Tecniche

Linfonodi locoregionali

Protesi descrizione

Terapie/follow-up

Annotation of radiology reports

File Options Tools Help

GATE

Applications

Language Resources

- GATE document_0001F
- Annotation schema_000
- Annotation schema_000
- Annotation schema_000
- Annotation schema_000
- Annotation schema_000
- Annotation schema_000
- Annotation schema_000
- Annotation schema_000
- Annotation schema_000
- Annotation schema_000
- Annotation schema_000

Processing Resources

Data stores

Messages GATE document_0001F

Annotation Sets Annotations Co-reference Editor Text

The Fall and Rise of Hugo Chavez

In Venezuela, the dust is finally settling after the dramatic turbulence of the past two weeks. **President Hugo Chavez** is once again at the helm of power, and the military has returned to its barracks. **But there are many grieving Venezuelans who aren't quite at peace.** Their loved ones were gunned down, first by sympathizers of Mr. Chavez and subsequently by **supporters** of the short-lived Interim government.

Mr. Chavez was ousted in a military coup on April 12 after 17 persons were killed during an **anti-Chavez protest**, apparently by **sharpshooters faithful to the president**. Pedro Carmona, a powerful businessman, stepped into power on Friday but was forced to resign the next day. Following the coup, **40 persons**, mostly Chavez supporters, were killed while protesting.

A probe of what transpired should be in the offing. **Mr. Chavez** has placed his cronies in every democratic institution, from congress to the courts, and **is expected to prevent any honest**

Check

- MPQA
- agent
- attitude
- expressive-su
- inside
- on
- split
- target

Type	Set	Start	End	
agent	MPQA	0	0	{id=implicit}
agent	MPQA	0	0	{id=w}
agent	MPQA	217	238	{id=chavez}
attitude	MPQA	324	390	{attitude-neg-intensity=medium, attitude-neg-toward=n}
agent	MPQA	338	363	{id=gv, nested-source=w,gv}
attitude	MPQA	343	351	{attitude-neg-intensity=medium-high, attitude-neg-tow}
attitude	MPQA	368	390	{attitude-neg-intensity=medium, attitude-neg-toward=n}
attitude	MPQA	392	425	{attitude-neg-intensity=high, attitude-neg-toward=none}

92 Annotations (0 selected)

Document Editor Initialisation Parameters

Edit gate options

Opinion Annotation in GATE

Facts, Sentiments and Big Data

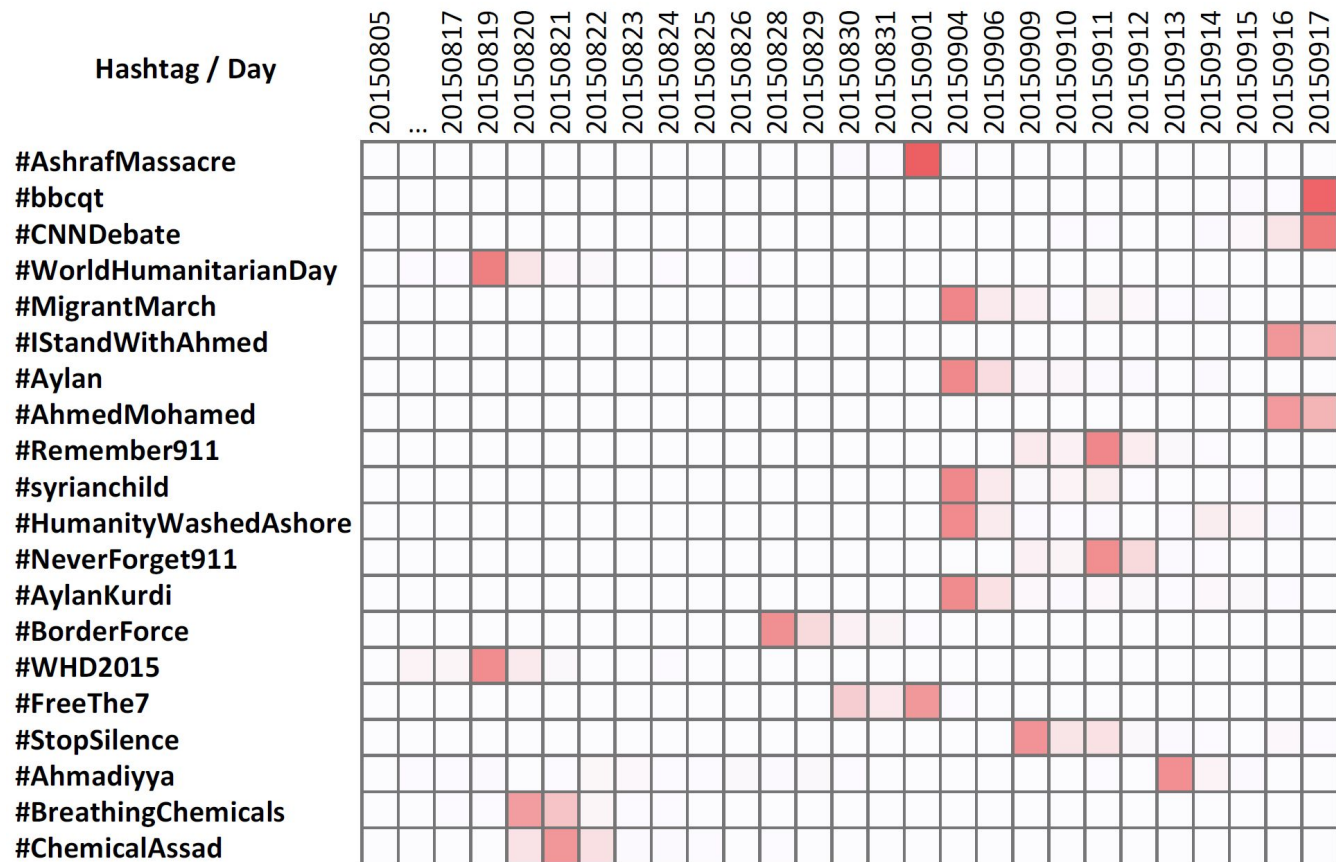
Facts and Big Data

When looking for factual information, the comparison of many sources of information allows to check for its truth, consistency and relevance.



Temporal/spatial anomalies in the use of language, e.g., spikes in the use of words like “earthquake”, “shots”, “explosion”, may allow to recognize events, and gather relevant data about them.

[Image source](#)



Event recognition from hashtag distribution over time

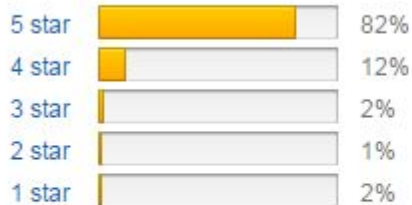
Sentiment and Big Data

Subjective information is varied by definition.

The more sources are compared, the more the vision of the feelings on the matter is complete.



4.7 out of 5 stars




[See all 1,771 reviews](#)

“It takes great photos and it is a very good quality camera.”

 | 478 reviewers made a similar statement

“Very easy to use as a point and shoot camera, as I will be taking some classes to really learn how to use more features.”

 | 362 reviewers made a similar statement

“It is the best thing I have spent money on, and I don't think anyone will regret buying this camera.”

 | 151 reviewers made a similar statement

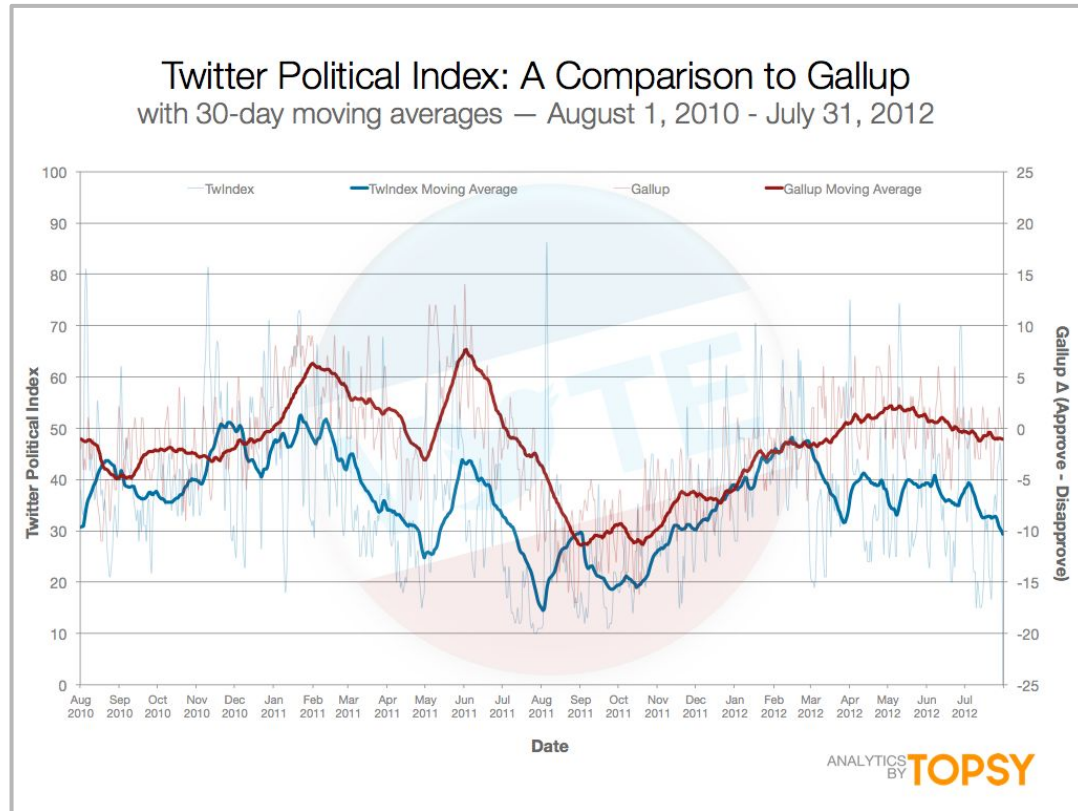
★☆☆☆☆ **Broke right out of the box**

By  on January 22, 2015

Style: w/ 18-55mm lens | Package Type: Standard Packaging

I have been begging for this camera for ever so I bought it right here off amazon as I got the the box I opened it and the hole box fell apart the camera fell and the lens focus ring broke as well as the lcd screen got scratch. I wouldn't be complaining but the camera didn't come with a warranty so I wasted 500 dollars on a broken camera.

Sentiment and Big Data



Why Sentiment Analysis?

(Is it of practical use?)

Why Sentiment Analysis?

When we have to take a decision we look for the opinion of the others.

The textual user-generated content that is

- shared on the Web/social networks,
- written in open-ended questions in questionnaires,
- sent to companies as feedback, . . .

contains

- **voluntarily produced,**
- **unconstrained,**
- **first-hand/personal,**
- **fresh,**

evaluative information about our topic of interest.

Why Sentiment Analysis?

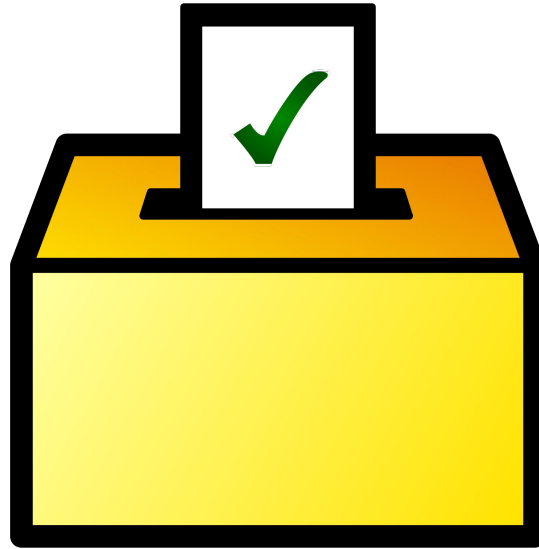
Practical example: customers satisfaction questionnaires.

- Are you happy with us? yes/no
- How much are you happy on a scale from 0 to 10?
- Your vote is determined by our: rates service other
- Write here any other feedback: _____

The first three answers can be directly **automatically processed** to extract **statistical information**.

The last answer to an **open-ended question** is the only potential source of **unexpected information**.

Why Sentiment Analysis?



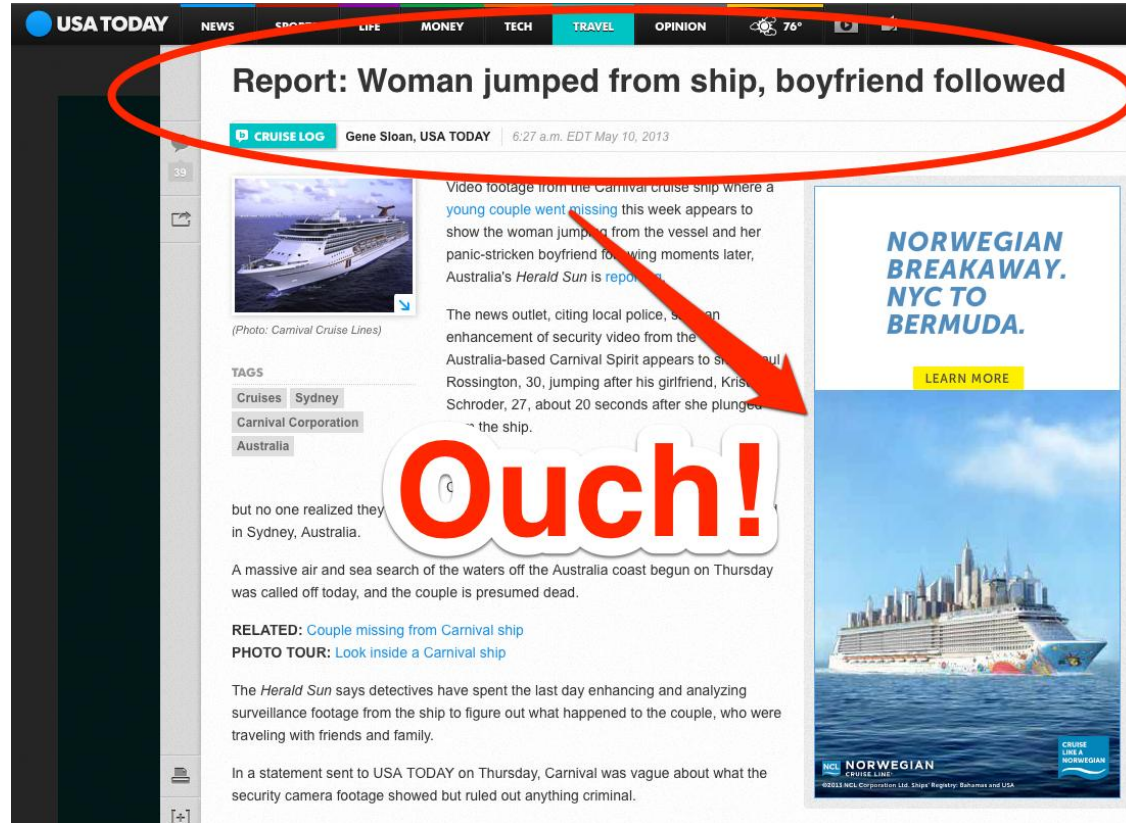
Why Sentiment Analysis?



Why Sentiment Analysis?




Why Sentiment Analysis?



USA TODAY NEWS SPORTS LIFE MONEY TECH TRAVEL OPINION 76°

Report: Woman jumped from ship, boyfriend followed

CRUISE LOG Gene Sloan, USA TODAY | 6:27 a.m. EDT May 10, 2013



(Photo: Carnival Cruise Lines)

Video footage from the Carnival cruise ship where a [young couple went missing](#) this week appears to show the woman jumping from the vessel and her panic-stricken boyfriend following moments later, Australia's *Herald Sun* is reporting.

The news outlet, citing local police, says an enhancement of security video from the Australia-based Carnival Spirit appears to show Paul Rossington, 30, jumping after his girlfriend, Kris Schroder, 27, about 20 seconds after she plunged over the ship.

but no one realized they were on the ship in Sydney, Australia.

A massive air and sea search of the waters off the Australia coast begun on Thursday was called off today, and the couple is presumed dead.


RELATED: [Couple missing from Carnival ship](#)
PHOTO TOUR: [Look inside a Carnival ship](#)

The *Herald Sun* says detectives have spent the last day enhancing and analyzing surveillance footage from the ship to figure out what happened to the couple, who were traveling with friends and family.

In a statement sent to USA TODAY on Thursday, Carnival was vague about what the security camera footage showed but ruled out anything criminal.

NORWEGIAN BREAKAWAY. NYC TO BERMUDA.

LEARN MORE



NCL NORWEGIAN CRUISE LINE
©2013 NCL Corporation Ltd. Ship Registry: Bahamas and USA

Why Sentiment Analysis?

→ Home > World > Article

Cambodian plane crashes, 20 feared dead

June 25, 2007 - 7:01PM

Tools sponsored by
Credit Cards     

A chartered plane flying today between two popular tourist destinations in Cambodia has crashed, with at least 20 people on board feared killed, an aviation official said.

The plane, a Russian-made AN-24, was flying from Siem Reap - where the famous Angkor Wat temple complex is located - to Sihanoukville, a coastal city with access to beaches, said Him Sarun, Cabinet chief for the Secretariat of Civil Aviation.

An official at Siem Reap airport said 13 of the passengers were from South Korea, three were Czech, one was Russian and five were Cambodian.

He said the plane carried a crew of five Cambodians.

The plane belonged to a small Cambodian

Advertisement

You fly.
We pay.



Over June 2007
we pay for road costs

© Scullin Advertising

Why Sentiment Analysis?

SFGate.com Quick Search GO

SFGATE HOME • BUSINESS • SPORTS • ENTERTAINMENT • TRAVEL JOBS • REAL ESTATE • AUTO



I AM TIGER. HEAR ME ROAR. PLAYING THE BAY AREA AUG. 17 – SEPT. 5! Get your Tickets Today!

AP Breaking News

Tiger Kills Teen During Senior Photo Shoot

Thursday, August 18, 2005

(08-18) 14:29 PDT Mound Valley, Kan. (AP) --

- [Printable Version](#)
- [Email This Article](#)

A Bengal tiger attacked and killed a teenage girl who was posing for her senior high school pictures at an animal sanctuary Thursday in southeast Kansas, authorities said.

The Labette County Sheriff's office identified the victim as Haley R. Hilderbrand, 17, of Altamont. A news release said Hilderbrand was at the Lost Creek Animal Sanctuary posing for a photo with the 7-year-old tiger, which was being restrained by its handler, when the animal attacked her.

Officers and handlers killed the animal. Emergency personnel were not able to revive Hilderbrand.



Beat the GMAT[®] price increase!

Enroll in our new comprehensive GMAT program by September 8, 2005 and lock in the current price.

KAPLAN
Test Prep and Admissions

Higher scores guaranteed or your money back.**

Learn more

* Test names are registered trademarks of their respective owners.
** Visit kaplan.com/frag for details.



on sale

Harveys Lake Tahoe Summer Concert Series
Harveys Lake Tahoe Summer Concert Series is underway!

Harvey's Lake Tahoe Bay Area

Space Saving Sleeping



EQUIFAX



Why Sentiment Analysis?

San Francisco Chronicle STORE
FOR THE VINO LOVER IN YOU! FREE SHIPPING



CHP: Drunk Aptos driver crashes into "Report Drunk Drivers" sign on Highway 1

Amy Larson, KSBW Published 9:00 am, Thursday, August 17, 2017

✉ f t p s + 📄 18



Photo: California Highway Patrol

IMAGE 1 OF 3

The driver, identified by the CHP as Stephen DeWitt, 57, of Aptos, was "quite intoxicated," one officer said.

Why Sentiment Analysis?

Q&A Here's why experts say all kids ages 6 and up should be screened for obesity



A government panel recommends that all children ages 6 and up be screened for obesity and referred for treatment if necessary. (Dreamstime)

By **Karen Kaplan**
Contact Reporter

JUNE 20, 2017, 3:55 PM

Right now,
a hungry
family
needs you.

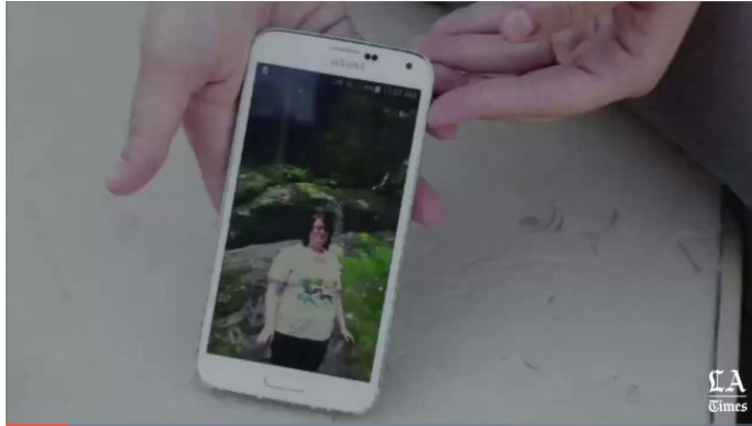
Donate food today!

WFP
World Food
Programme
wfp.org

ADVERTISEMENT

Why Sentiment Analysis?

As obesity keeps rising, more Americans are just giving up



New research shows if you know you're overweight or obese, and you know your extra pounds are unhealthy, that you may not be making a stab at losing weight anymore.



By **Melissa Healy** · Contact Reporter

MARCH 7, 2017, 3:35 PM

TIt stands to reason that if you know you're overweight or obese, and you know your extra pounds are unhealthy, that you've made a stab at losing weight. Right?

Right now,
a hungry
family
needs you.

Donate food today!

WFP
World Food Programme
wfp.org

ADVERTISEMENT

Why Sentiment Analysis?

Yahoo! News Sign In
New User? Sign Up

Home U.S. Business World Entertainment Sports Tech Politics Science Health Travel Most Popular

Most Unread Most Viewed Most Recommended

Search: [] All News & Blogs Search: []

Most Popular Select a Category: Most Popular

MOST POPULAR HIGHLIGHT

Coffee Might Trigger First Heart Attack in Some

Updated 2 hours, 3 minute ago

TUESDAY, Aug. 15 (HealthDay News) — An occasional cup of coffee might trigger first heart attacks in some people, a new study suggests.

[View related sites](#) | [Include statistics](#) | [Include photos](#)

Most Emailed News [NY 11](#) [RSS](#)

-  **Home sales decline in 28 states, D.C.**
Updated 2 hours, 42 minutes ago
Sent 3,188 times
WASHINGTON — The slowdown in the once-sizzling housing market is spreading, with 28 states and the District of Columbia reporting spring sales declines, led by big drops in former boom areas of Arizona.

ADVERTISEMENT



Enjoy a Second Cup

[Learn](#) [Watch](#) [Play](#) [Subscribe](#)

MOST EMAILED PHOTOS [NY 11](#)

Why Sentiment Analysis?



Health News

Share this article:



Print



Study: 75 percent of Americans will be overweight by 2020

Associated Press

Posted on September 23, 2010 at 12:41 PM



Recommend

PARIS – Citizens of the world's richest countries are getting fatter and fatter and the United States is leading the charge, an organization of leading economies said Thursday in its first ever obesity forecast.


Three out of four Americans will be overweight or obese by 2020, and disease rates and health care spending will balloon, unless governments, individuals and industry cooperate on a comprehensive strategy to combat the epidemic, the study by the Organization for Economic Cooperation and Development said.

Why Sentiment Analysis?

11-year-old charged with driving drunk

REUTERS 

NEWS ALERTS

Get an alert when there are new stories about 

- Orange Beach, Alabama
- Perdido Key, Florida
- Chevrolet Monte Carlo

Add Selected Alerts

» More Alerts

ELSEWHERE ON THE WEB

CNN.COM

Explosion rocks besieged mosque in Pakistani capital

ABC NEWS

'Out of the Blue': Do Aliens Exist?

THE CHRISTIAN SCIENCE MONITOR

William Bratton: Lauded chief of troubled LAPD



Fri Jul 6, 3:23 PM ET

MIAMI (Reuters) - An 11-year-old girl was charged with drunken driving after leading police on a chase at speeds of up to 100 mph that ended when she flipped the car in an Alabama beach town.

A video camera in the police car captured the look of surprise on the officer's face when he approached the wrecked car and got a look at the motorist.

The Mobile Press-Register newspaper said the patrolman saw the Chevrolet Monte Carlo speeding and flashed his lights to signal the driver to stop. Instead, the car sped faster, traveling at up to 100 mph (160 kph) before sideswiping another vehicle and flipping over in the Gulf Coast town of Orange Beach, Alabama, on Tuesday night.

The young driver, who lived nearby in Perdido Key, Florida, was treated at a hospital for scrapes and bruises and released to relatives. Police also charged her with speeding, leaving the scene of an accident and reckless endangerment.

The car belonged to a relative and police were still trying to find out where she got the alcohol. There was none in the vehicle but her blood alcohol level was over the limit for adult motorists, police told the newspaper.

[Email Story](#) [IM Story](#) [Printable View](#)

ADVERTISEMENT

HOORAY BEER!

Click here for more BOO

IMPORTED
Pilsener
LAGER

Please drink and drive responsibly.
HOORAY BEER! and associated signs and letter-
marks are trademarks of Carlsberg & Carlsberg Limited. HOORAY
Lager ©2011 Carlsberg & Carlsberg, imported by Carlsberg - America USA, Torrance, CA.

Why Sentiment Analysis?



Hostess fa sesso con i passeggeri: "Ho guadagnato circa 900mila euro in due anni"



Attualità

Cinema

Musica

Ritratti

Cultura

Stile

Televisione

Cucina

Donne

La Musica è

Lavoro

Segui P@ Magazine su:



Why Sentiment Analysis?

San Francisco Chronicle

World

AMAZON WEDDING REGISTRY

All you.
One registry.

Create your registry now

Privacy

Instant divorce unconstitutional, India's top court rules

By Shashank Bengali and Parth M.N. | August 22, 2017

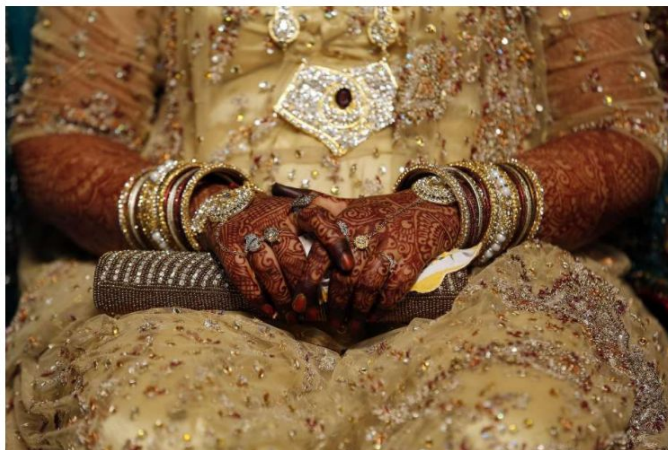


Photo: Rajanish Kakade, Associated Press



A Muslim bride participates in a mass wedding in 2014 in Mumbai. India is home to nearly 200 million Muslims — the largest minority in a Hindu-majority country of 1.3 billion people.

Sentiment Analysis tasks

Sentiment Analysis tasks

Most of SA research and applications are focused on the simple **positive vs negative** dichotomy (or a graded scale among this two opposites).

Most common SA tasks:

- Subjectivity/polarity classification
- Regression
- Opinion extraction
- Quantification

There is also research on **emotions**, **attitude** and **humor** in human language.

Classification

Classification: determining the attitude of the author of a document toward the document subject matter.

By **subjectivity**: determining if the text contains or not subjective evaluations.

“The movie is set in WW2” → Objective

“The plot is confusing” → Subjective

By **polarity**: determining if the subjective evaluations are positive or negative with respect to its topic.

“This movie is a masterpiece” → Positive

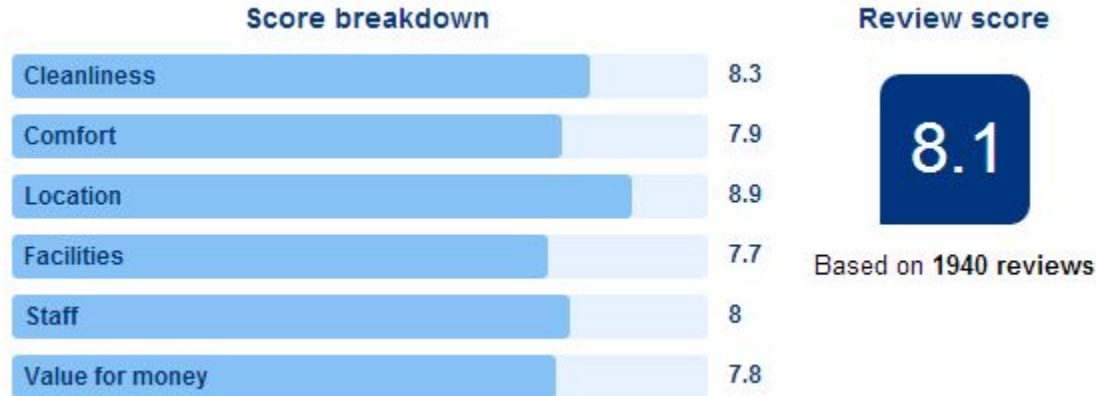
Regression

Regression: extending the polarity classification problem to a ordinal scale.

Typical scenario: "Star rating" of product reviews.

"This phone is not worth its price" → 

Regression can produce a global evaluation or be focused on specific aspects.



Extraction

Extraction: identifying the expressions of an opinion, its properties, and the target of that opinion.

*“The phone has a **great display** but it is killed by the **small battery**”*

(**display**: **great**, **positive**), (**battery**: **small**, **negative**)

Extraction is often modeled as a **classification** problem **at the word level**.

The output of extraction contribute to build a **knowledge base**, which can be then **queried** by traditional methods from Information Retrieval and **Data Mining**.

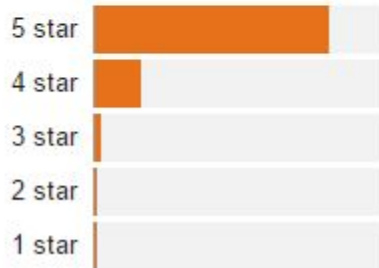
Extraction



Apple iPhone 6 - 16 GB - Space Gray - Unlocked - CDMA/GSM

\$380 online ★★★★★ 7,092 product reviews

Reviews



4.8



7,092 reviews

size

battery

ease of use

design/style

value

picture/video

customer service

"Overall happy with the phone, like the size most."

"Overall a good phone and wish a better battery life"

"Great phone easy to use."

"Overall phone is good in design and weight."

"Great overall item for the price would reccomend it!"

"Very fast, awesome picture quality, user friendly."

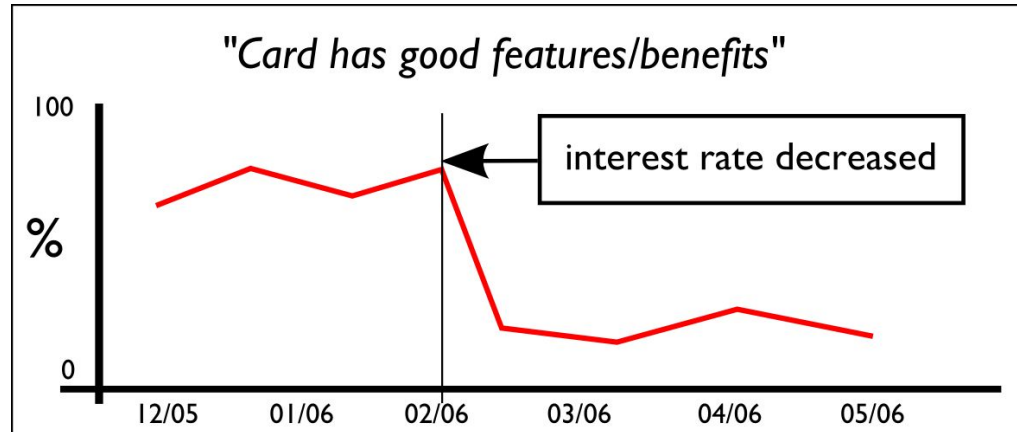
"Best Buy has great customer service."

Example of extraction of aspect-related relevant evaluations, Google Shopping

Quantification

Quantification is a problem of aggregate analysis: a set of documents is processed as a single entity in order to determine some properties of the whole set.

- Determining the proportion, and its trend over time, of positive reviews about a product.



Sentiment Analysis methods

There is no one-stop solution for Sentiment Analysis.

Sentiment Analysis is not a single problem.

Sentiment Analysis is not a dataset.

Sentiment Analysis is not a lexicon.

Sentiment Analysis is not an algorithm.

Sentiment Analysis is a special scenario for text analysis problems.

A “standard” method produces 70-90% of the result.

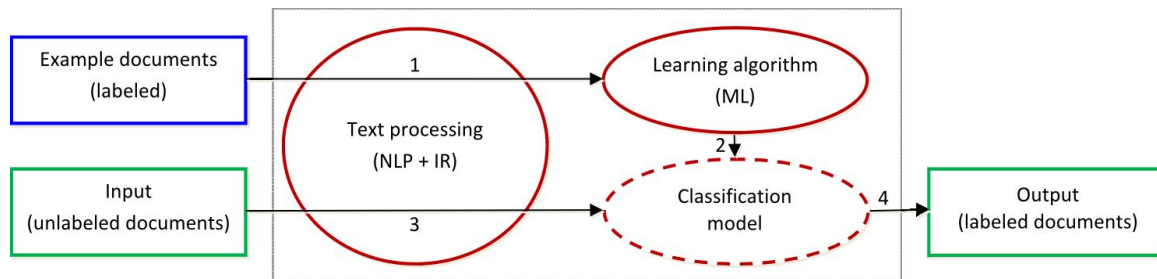
Exploiting the characteristic that are specific of a given Sentiment Analysis problem produces that 10-30% improvement that separates an average solution from a good one.

Sentiment Analysis methods

Multidisciplinary approach:

- Natural Language Processing
- Information Retrieval
- Machine Learning

The template solution to a sentiment analysis problem is the same of a “generic” one, e.g.:



Most of sentiment-specific methods deal with **capturing how sentiment are expressed in natural language.**

The language of opinions

The language of opinions

The language we use to express our subjective evaluations is **one of the most complex parts of language**.

There are many components in the language of opinions:

- Global/Domain-specific lexicon.
- Valence shifters/Comparative expressions.
- Irony, sarcasm, common knowledge.
- Other aspects, e.g., morphology. . .

The main aim of NLP/IR/ML applied to Sentiment Analysis is to **recognize** sentiment expressions and to **model** them into **semantic abstractions**.

The language of opinions

Some **words** have a **globally** recognized **sentiment valence** in any context of use, e.g.: “good”, “poor”, “perfect”, “ugly”

“A good tool that works perfectly”

“I had an horrible experience”

General purpose lexical resources list these words associating sentiment labels to them, e.g.:

- The General Inquirer lexicon
- WordNet affect
- SentiWordNet

The language of opinions

Domain/aspect-specific expressions: words that have a sentiment valence only when used in the context of a **specific domain**, or when they are associated with a **specific aspect**.

*“The phone is made of **cheap** plastic”*

*“The carrier offers **cheap** rates”*

*“We have got a **warm** welcome”*

*“We have got a **warm** beer”*

A collection of text from the domain can be used to build a **domain lexicon**.

The language of opinions

Negation and **valence shifters**: they do not determine sentiment directly but have **influence** on it.

It is difficult to determine their **scope** and **combined effect**.

*“This is a **very good** car” (increment)*

*“This car is **not very good**” (flip, decrement)*

*“I **don't like** the design of the new Nokia **but** it contains some **intriguing** functions”*

*“**Not** only is this phone **expensive** **but** it is also **heavy** and **difficult** to use”*

[Workshop on Negation and Speculation in NLP](#)

The language of opinions

Punctuation, emoticons, emoji:

"7AM battery 100% - 9AM 30% 😞"

Irony, sarcasm:

"Light as a bulldozer"

"The most useful idea since the DVD rewinder"



Common knowledge:

"Windows Vista: the new Windows ME"

"Windows 7: the new Windows XP"



Sentiments, Emotions, Humor

Affective computing

Modern Sentiment Analysis applications are mainly **data mining oriented** and focused on the evaluations expressed toward the subject matter of the text.

There is also active research on the topic of **affective computing**, more related to **psychology** and **cognitive sciences**.

In affective computing the focus is on the **human computer interaction**, aiming at identifying the **emotions** and **feelings conveyed** by the text **to the reader**.

Affective computing

Recognizing the expression of six basic emotions: anger, disgust, fear, joy, sadness and surprise:

“He looked at his father lying drunk on the floor” (disgust)

“She was leaving and she would never see him again” (sadness)

“She turned and suddenly disappeared from their view” (surprise)

“They celebrated their achievement with an epic party” (joy)

[Strapparava and Mihalcea. Learning to Identify Emotions in Text. SAC 2008](#)

Computational humor

Generating and recognizing humor: jokes, puns, wordplay.

"Beauty is in the eye of the beholder"

"Beauty is in the eye of the beer holder"

Generation is usually based on templates, recognition is mainly based on stylistic features.

An example of application is building a language playground for people with complex communication needs.

[Ritchie et al. *A practical application of computational humour*. ICCV 2007.](#)

[Mihalcea and Strapparava. *Learning to Laugh \(Automatically\): Computational Models for Humor Recognition*. Computational Intelligence, 2006.](#)

Irony and sarcasm

Irony and sarcasm are pervasive on social media.



Both are linguistic phenomena that rely on context and common knowledge.

 **Donald J. Trump** 
@realDonaldTrump [Segui](#) 


President Obama played golf yesterday???


 Traduci dalla lingua originale: inglese

13:59 - 18 nov 2013

 [Segui](#) 

"President Obama played golf yesterday???"

 Traduci dalla lingua originale: inglese



10:54 - 30 giu 2017

Irony and sarcasm

Research on computational recognition of irony is at an early stage, mainly focusing on syntactic features.

Data is often collected from tweets with *#ironic* or *#sarcasm* hashtag.



[Wallace, "Computational irony: A survey and new perspectives" AIR 2015](#)
[Hernández & Rosso "Irony, Sarcasm, and Sentiment Analysis" Chapter 7 in](#)
["Sentiment Analysis in Social Networks" Liu, Messina, Fersini, Pozzi](#)