### 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 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.

**Objective information:** 

### The **4.7-inch** <u>display</u> on the iPhone 6 is arguably its best feature.

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



Subjective information:

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

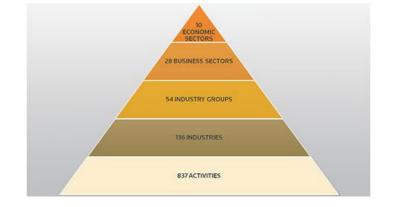
...concerns have been raised about the **relatively low** <u>resolution</u> (1334 x 750 pixels)



Classification of documents:

• with respect to the Thomson Reuters

taxonomy\*.



 with respect to the content being a positive, neutral, or a negative evaluation<sup>†</sup>.



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 <mark>generally satisfied</mark> with termination packages being offered<sup>†</sup>

Agent <mark>Attitude</mark> Target

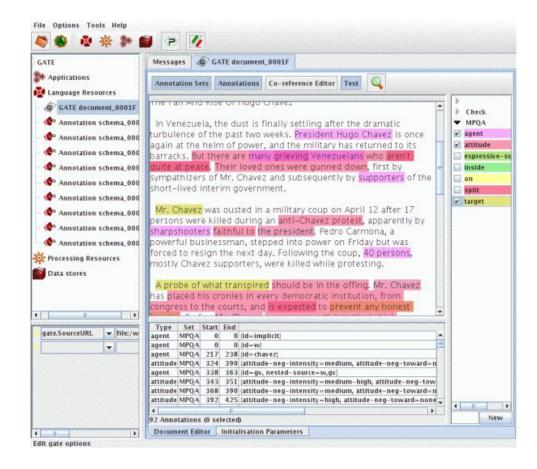


L'esame È stato eseguito con sequenza T2 STIR e con sequenze T1 3D dinamiche prima e dopo somministrazione di mdc paramagnetico, acquisite secondo piani di scansione assiali. Diffusi esiti cicatriziali in sede retroareolare sinistra. Non si apprezzano potenziamenti sospetti bilateralmente. In particolare non si È osservato un corrispettivo RM dell'immagine descritta mammograficamente a sinistra. In sede ascellare bilateralmente si apprezzano alcune linfoadenopatie 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/assen	za
Indicazioni Esame	
Informazioni Tecniche	
Linfonodi locoregionali	
Protesi descrizione	
Terapie/follow-up	

### Annotation of radiology reports

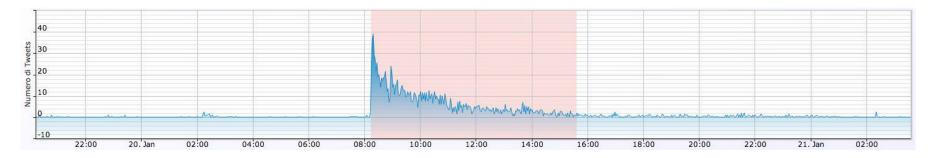


Opinion Annotation in <u>GATE</u>

### 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.



Hashtag / Day	20150805		015081	015081	015082	015082	20150822	20150823	20150824	508	5082	508	20150829	20150830	$\circ$	20150904	$\mathbf{O}$	20150909	20150910	20150911	20150913	20150914	-	20150916	20150917
#AshrafMassacre																									
#bbcqt																									
#CNNDebate																									
#WorldHumanitarianDay																									
#MigrantMarch																									
#IStandWithAhmed																									
#Aylan	Ц	$ \downarrow$		$ \perp$	$\downarrow$																				
#AhmedMohamed	Ц	$ \downarrow$			$\downarrow$																				
#Remember911	Ц	$ \rightarrow$		$ \rightarrow$	$\downarrow$	$\downarrow$																			
#syrianchild	$\square$	$ \downarrow$			$\downarrow$		_																		
#HumanityWashedAshore	$\square$	$ \rightarrow$	4	$\rightarrow$	$\downarrow$	$\downarrow$	_																		Ц
#NeverForget911	$\square$	$\rightarrow$	4	$\rightarrow$	$\downarrow$	$\downarrow$	_										_							$\square$	$\square$
#AylanKurdi	$\square$	$\rightarrow$	_	$\rightarrow$	$\rightarrow$	$\rightarrow$																		$\square$	$\square$
#BorderForce	$\square$	$\rightarrow$			4		_																	$\square$	$\square$
#WHD2015	$\square$				_	4	_										_								$\square$
#FreeThe7	$\square$	$\rightarrow$	+	$\rightarrow$	+	$\dashv$	_										_						_		$\square$
#StopSilence	$\square$	$\rightarrow$	_	$\rightarrow$	+	$\downarrow$	_										_						_		$\square$
#Ahmadiyya	$\square$	$\rightarrow$	+	_			_																		$\square$
#BreathingChemicals		$\rightarrow$	_		4		_										_								
#ChemicalAssad																									

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 c	of 5 stars
5 star	82%
4 star	12%
3 star	2%
2 star	1%
1 star	2%

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."

"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

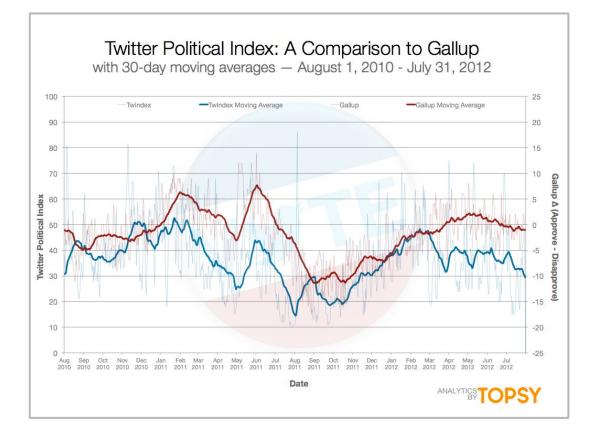
### ATTTTT Broke right out of the box

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



**Twindex** 

# Why Sentiment Analysis? (Is it of practical use?)

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,

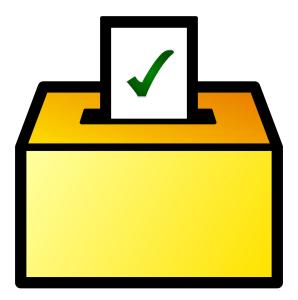
evaluative information about our topic of interest.

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:  $\Box$  rates  $\Box$  service  $\Box$  other
- Write here any other feedback: \_\_\_\_\_

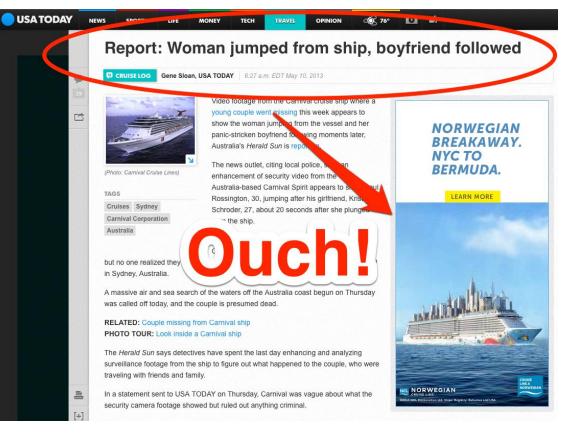
The first three answers can be directly **automatically processed** to extract **statistical information**.

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









-+ Home > World > Article

### Cambodian plane crashes, 20 feared dead

June 25, 2007 - 7:01PM

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.

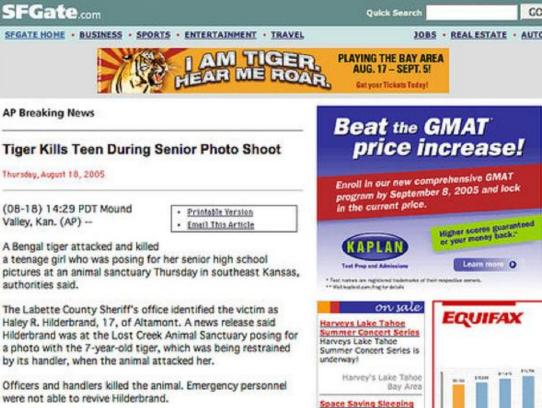
An official at Siem Reap airport said 13 of the passengers were from South Korea, three were Czech, one was Russian and five were Cambodian.

He said the plane carried a crew of five Cambodians.

The plane belonged to a small Cambodian

Advertisement You fly. We pay. UNTER USA 5270 we pay on wad bosts

Tools sponsored by Credit Cards Control Day



A Bengal tiger attacked and killed

pictures at an animal sanctuary Thursday in southeast Kansas, authorities said.

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.

were not able to revive Hilderbrand.

SFGATE LOCAL NEWS SPORTS BUSINESS A&E FOOD LIVING TRAVEL REALESTATE (



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

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



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.

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





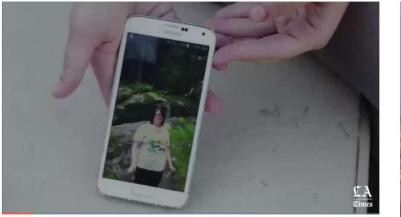
ADVERTISEMENT

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

### 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

t 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.







Health News



### 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.

### 11-year-old charged with driving drunk

town.

Fri Jul 6, 3:23 PM ET

### REUTERS

### NEWS ALERTS

Get an alert when there are new stories about:

Corange Beach, Alabama

Perdido Key, Florida

Chevrolet Monte Carlo

### Add Selected Alerts

More Alerts

### ELSEWHERE ON THE WEB

CNN.COM Explosion rocks besleged mosque in Pakistani capital

ABC NEWS 'Out of the Blue': Do Aliens Exist?

THE CHRISTIAN SCIENCE MONITOR William Bratton: Lauded chief of troubled LAPD 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

A video camera in the police car

wrecked car and got a look at the

captured the look of surprise on the officer's face when he approached the

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.

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

🖾 Email Story 斗 IM Story 📇 Printable View







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

Necessities got you in a pinch? Click to sign up for your chance to **win a \$200 gift card.** 

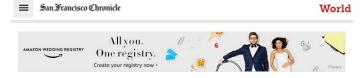


Attualità

Cinema Musica Ritratti Cultura Stile Televisione Cucina Donne La Musica è Lavoro

Segui FQ Magazine su:

Segui @FQMagazine



### Instant divorce unconstitutional, India's top court rules

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

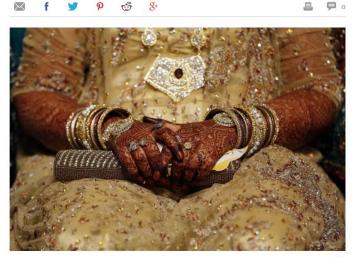


Photo: Rajanish Kakade, Associated Press

1

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  $WW2" \rightarrow Objective$ 

"The plot is confusing"  $\rightarrow$  Subjective

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

"This movie is a masterpiece"  $\rightarrow$  Positive

# Regression

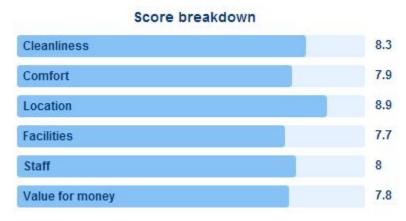
Regression: extending the polarity classification problem to a ordinal scale.

Typical scenario: "Star rating" of product reviews.

"This phone is not worth its price"

$$\bullet \bigstar \And \And \checkmark \checkmark$$

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



### **Review score**



#### 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

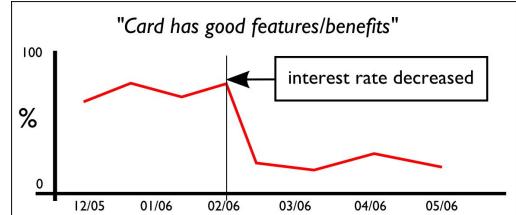


#### Example of extraction of aspect-related relevant evaluations, Google Shopping

# 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.

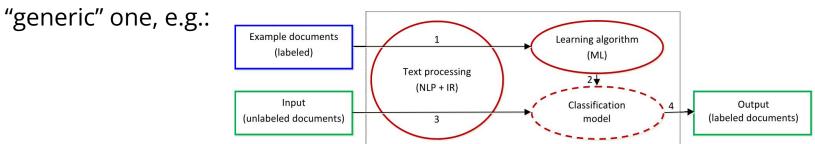
Exploiting the characteristic that are specific of a given Sentiment Analysis problem produces that 10-30% improvement that separates an average solution from a good one.

## 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



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

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**.

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

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**.

**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

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. ICCC 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.



### 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.





21:40 - 18 nov 2017

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