

Introduzione alla Sentiment Analysis - I

Rachele Sprugnoli – rachele.sprugnoli@unicatt.it

Centro Interdisciplinare di Ricerche per la Computerizzazione
dei Segni dell'Espressione (CIRCSE)



UNIVERSITÀ
CATTOLICA
del Sacro Cuore



COS' È LA 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

Tanti nomi, tanti task:

- opinion extraction
- sentiment mining
- subjectivity analysis
- affect analysis
- emotion analysis
- review mining
- ...



TESTI OGGETTIVI *versus* SOGGETTIVI

TESTO OGGETTIVO

«La Pinacoteca di Brera è una galleria nazionale d'arte antica e moderna»

- Racconta dei fatti
- Verificabile
- Presentato con obiettività

TESTO SOGGETTIVO

«Entrare in Pinacoteca significa essere travolti dalla bellezza»

- Racconta opinioni, valutazioni, emozioni, giudizi, punti di vista
- Non verificabile
- Presentato senza obiettività



PERCHÉ LE OPINIONI SONO IMPORTANTI

- Il contenuto generato dagli utenti è sempre maggiore anche:
 - nei social media e sul web
 - nelle comunicazioni dei consumatori/dipendenti alle aziende
 - nelle risposte alle domande aperte dei questionari

3) Giudizio complessivo sull'assistenza ed eventuali suggerimenti per migliorare il servizio

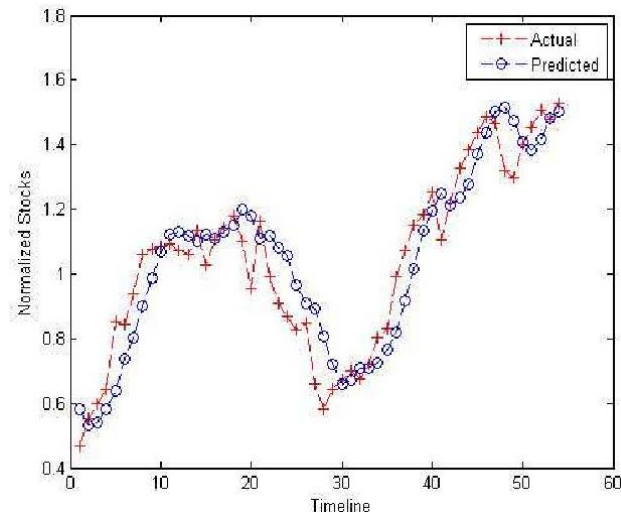
<input type="checkbox"/> Ottimo
<input type="checkbox"/> Buono
<input type="checkbox"/> Sufficiente
<input type="checkbox"/> Insufficiente
Se insufficiente, spiegare motivazioni:
Eventuali suggerimenti:



PERCHÉ LE OPINIONI SONO IMPORTANTI

- Applicazioni
 - ricerche di mercato
 - ricerche politiche
 - ricerche sociali
 - previsioni: risultati di elezioni politiche, andamento delle vendite o del mercato azionario, entrate al botteghino per i film...

Mittal and Goel, "Stock Prediction Using Twitter Sentiment Analysis", 2012



COSA SONO LE OPINIONI

- Sono formate da 5 elementi:



COSA SONO LE OPINIONI

- Dati non strutturati: il linguaggio delle opinioni

- **Prior Polarity:** la polarità delle singole parole considerate al di fuori di qualsiasi contesto

- *eccezionale* = POS
- *migliore* = POS
- *ottimo* = POS



visioneM
Milano, Italia
10 5



Recensito 2 settimane fa

Pizze napoletane eccezionale vale la lunga attesa

Il locale non accetta prenotazione quindi è molto probabile fare 50/60 minuti di attesa fuori dal locale in attesa di essere chiamati, ma l'attesa viene ripagata dalla soddisfazione di cenare nella migliore pizzeria di Milano. La pizza è napoletana doc, il menu è diviso tra pizze classiche e pizze regionali per le quali utilizzano ingredienti di altissima qualità. Il prezzo è ottimo rapportato alla qualità dei piatti.

Data della visita: febbraio 2020



Qualità/prezzo



Servizio



Cibo

Chiedi a visioneM in merito a Pizzium - Via Vigevano

Grazie, visioneM

Questa recensione rappresenta l'opinione personale di un viaggiatore di TripAdvisor e non di TripAdvisor LLC.



COSA SONO LE OPINIONI

- Dati non strutturati: il linguaggio delle opinioni

- **Semantic Orientation:** la misura della polarità e dell'intensità di un'opinione



visioneM
Milano, Italia
10 5

- *lungo + attesa* = **NEG**
- *dolce + attesa* = **POS**
- *lungo + tradizione* = **POS**
- *alto + qualità* = **POS**
- *basso + qualità* = **NEG**
- *alto + prezzo* = **NEG**



Recensito 2 settimane fa

Pizze napoletane eccezionale vale la lunga attesa

Il locale non accetta prenotazione quindi è molto probabile fare 50/60 minuti di attesa fuori dal locale in attesa di essere chiamati, ma l'attesa viene ripagata dalla soddisfazione di cenare nella migliore pizzeria di Milano. La pizza è napoletana doc, il menu è diviso tra pizze classiche e pizze regionali per le quali utilizzano ingredienti di altissima qualità. Il prezzo è ottimo rapportato alla qualità dei piatti.

Data della visita: febbraio 2020



Qualità/prezzo



Servizio



Cibo

Chiedi a visioneM in merito a Pizzium - Via Vigevano

Grazie, visioneM

Questa recensione rappresenta l'opinione personale di un viaggiatore di TripAdvisor e non di TripAdvisor LLC.



COSA SONO LE OPINIONI

- Dati non strutturati: il linguaggio delle opinioni

- **Negazioni e Sentiment shifters:** elementi e strutture linguistiche capaci di modificare la *prior polarity*
 - **NON** accetta prenotazioni...



visioneM
Milano, Italia
10 5



Recensito 2 settimane fa

Pizze napoletane eccezionale vale la lunga attesa

Il locale non accetta prenotazione quindi è molto probabile fare 50/60 minuti di attesa fuori dal locale in attesa di essere chiamati, ma l'attesa viene ripagata dalla soddisfazione di cenare nella migliore pizzeria di Milano. La pizza è napoletana doc, il menu è diviso tra pizze classiche e pizze regionali per le quali utilizzano ingredienti di altissima qualità. Il prezzo è ottimo rapportato alla qualità dei piatti.

Data della visita: febbraio 2020



Qualità/prezzo



Servizio



Cibo

Chiedi a visioneM in merito a Pizzium - Via Vigevano

Grazie, visioneM

Questa recensione rappresenta l'opinione personale di un viaggiatore di TripAdvisor e non di TripAdvisor LLC.



IL LINGUAGGIO DELLE OPINIONI

- Il linguaggio che usiamo per esprimere le nostre valutazioni soggettive è estremamente complesso, formato da molte componenti:

- lessico con un sentiment che cambia in base al contesto

«abbiamo avuto un caldo benvenuto»

«abbiamo avuto una birra calda»

«il telefono è fatto di plastica economica»

«la compagnia aerea offre tariffe economiche»



IL LINGUAGGIO DELLE OPINIONI

- Il linguaggio che usiamo per esprimere le nostre valutazioni soggettive è estremamente complesso, formato da molte componenti:

- ironia, sarcasmo, emoticon

«veloce come un bradipo» = NEG

«che telefono fantastico! Si è rotto dopo 2 giorni!» = NEG

«ore 7:30 batteria al 100% - ore 21 batteria 20% 😞» = NEG

«🔥🔥🔥» = POS

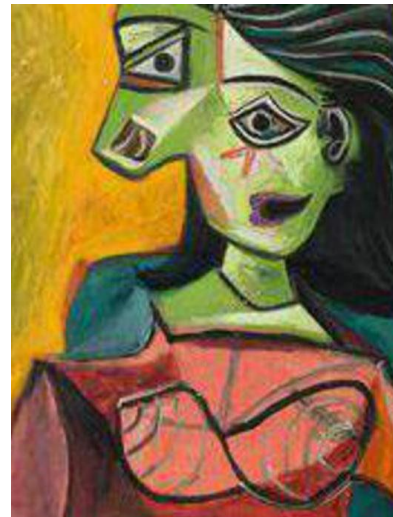


IL LINGUAGGIO DELLE OPINIONI

- Il linguaggio che usiamo per esprimere le nostre valutazioni soggettive è estremamente complesso, formato da molte componenti:
 - conoscenza del mondo

« Ha il viso come una madonna di Botticelli! » = POS

« Sembra un quadro di Picasso! » = NEG



IL LINGUAGGIO DELLE OPINIONI

- Il linguaggio che usiamo per esprimere le nostre valutazioni soggettive è estremamente complesso, formato da molte componenti:

- opinioni e sentiment impliciti

«Ho comprato il materasso una settimana fa e adesso c'è un avvallamento» versus «Il materasso si è rovinato dopo poco»

«La batteria degli iPhone dura di più di quella dei Samsung» versus «La batteria degli iPhone è migliore di quella dei Samsung»



TASK DI SENTIMENT ANALYSIS

- Non esiste una soluzione unica per la Sentiment Analysis
 - la Sentiment Analysis non è un singolo task
 - la Sentiment Analysis non è un singolo algoritmo
- Breve descrizione di task
 - Subjectivity detection
 - Document sentiment classification
 - Sentence-level sentiment classification
 - Aspect-based sentiment analysis
- Altri task strettamente correlati alla Sentiment Analysis
 - Stance detection
 - Irony detection
 - Emotion analysis
 - Hate speech/cyberbullying/misogyny detection



1) SUBJECTIVITY DETECTION

- Determinare se un testo contiene o meno delle valutazioni soggettive

Oggettivo

pisa93

domenica 16
febbraio 2014

un film che ha ben poco di monumentale ★★☆☆☆

Durante la Seconda Guerra Mondiale, un manipolo di uomini chiamati Monuments Men viene incaricato di recuperare la maggior parte delle opere d'arte trafugate dai Tedeschi. Tratto da una storia vera. Giunto alla sua quinta regia George Clooney vorrebbe stupire con un film alla Spielberg, ma non riesce ad essere incisivo. Si ispira alle ambientazioni di "Salvate il soldato Ryan" ed all'introspezione di "La sottile linea rossa", ma tutto ciò che ottiene è un involucro vuoto, senz'anima. Monumets Men è un film che fatica a trovare una sua identità, alternando commedia e dramma con la stessa semplicità con cui una ragazza si cambia d'abito. Manca proprio di quella profondità di cui un film del genere dovrebbe essere permeato. [+]

[+] lascia un commento a pisa93 »



Soggettivo



2) DOCUMENT SENTIMENT CLASSIFICATION

- Classificare un intero documento in base al sentiment:
 - il documento esprime un'opinione su una singola entità da parte di un singolo «holder»
 - scelta delle classi: categorie o scala di valori numerici?

NEGATIVE? VERY NEGATIVE? -2?

pisa93

un film che ha ben poco di monumentale ★★☆☆☆

Durante la Seconda Guerra Mondiale, un manipolo di uomini chiamati Monuments Men viene incaricato di recuperare la maggior parte delle opere d'arte trafugate dai Tedeschi. Tratto da una storia vera. Giunto alla sua quinta regia George Clooney vorrebbe stupire con un film alla Spielberg, ma non riesce ad essere incisivo. Si ispira alle ambientazioni di "Salvate il soldato Ryan" ed all'introspezione di "La sottile linea rossa", ma tutto ciò che ottiene è un involucro vuoto, senz'anima. Monuments Men è un film che fatica a trovare una sua identità, alternando commedia e dramma con la stessa semplicità con cui una ragazza si cambia d'abito. Manca proprio di quella profondità di cui un film del genere dovrebbe essere permeato. **[+]**

[+] lascia un commento a pisa93 »



3) SENTENCE-LEVEL SENTIMENT CLASSIFICATION

- Classificare una frase in base al sentiment:
 - un documento può contenere opinioni anche opposte
 - vari livelli di granularità possibili: paragrafi, strofe...

no_data

l'arte di essere americani ★★☆☆☆

L'idea di fondo non è male: l'arte come patrimonio, come radice della nostra identità, del nostro essere; la cultura come punto di partenza per ricordarci che siamo umani e, da lì, per ricostruire un continente distrutto dalla guerra. Questa idea merita rispetto e bisogna riconoscere come sia effettivamente vera. Tuttavia, il film in alcuni punti sembra perdere di vista questa linea: purtroppo spesso sembra più impegnato a dimostrare quanto sono stati bravi e meticolosi gli Americani e quanto di buono abbiano fatto in Europa. La realtà è, come spesso accade, molto più complessa! Gli episodi in cui le forze alleate hanno distrutto opere d'arte di grande valore durante il secondo conflitto mondiale sono noti e il film forse avrebbe dovuto fare un pò di autocritica in più. [+]

[+] lascia un commento a no_data »

martedì 29 luglio 2014

Si No
0% 0%
d'accordo?

POSITIVE

NEGATIVE



4) ASPECT-BASED SENTIMENT ANALYSIS

- Individuare le singole caratteristiche che rappresentano i punti di forza e/o le debolezze discusse nell'opinione
 - chiamato anche *feature-based sentiment analysis*
 - 2 fasi principali:
 - 1) estrazione delle caratteristiche
 - 2) classificazione del sentiment per ciascuna caratteristica



fred02



Carino, da provare.

Recensito in Italia il 2 novembre 2018

Style: Echo Dot | Colore: Antracite | Configurazione: Solo dispositivo

Ingolosito dal prezzo scontato ho acquistato Alexa e mi è puntualmente arrivato il 30 ottobre. Ho avuto difficoltà a configurarlo perché l'app. Alexa, sul mio OP5 non permetteva la configurazione. Questo è un problema che sembra affliggere anche i dispositivi Xiaomi. L'ho poi configurato con un Ipad. Carino, l'ho paragonato al Google Home mini che già possiedo, e sinceramente preferisco quest'ultimo. Comunque prodotto valido e divertente, al prezzo di lancio è sicuramente da provare.

7 persone l'hanno trovato utile



4) ASPECT-BASED SENTIMENT ANALYSIS

- 2 fasi principali:
 - 1) **estrazione** delle caratteristiche
 - 2) classificazione del sentiment per ciascuna caratteristica

CONFIGURAZIONE

PREZZO

ESTETICA

The image shows a screenshot of an Amazon product review for the Echo Dot. The review is by user 'fred02' and is a 4-star rating. The text of the review is: 'Ingoloso dal prezzo scontato ho acquistato Alexa e mi è puntualmente arrivato il 30 ottobre. Ho avuto difficoltà a configurarlo perché l'app. Alexa, sul mio OP5 non permetteva la configurazione. Questo è un problema che sembra affliggere anche i dispositivi Xiaomi. L'ho poi configurato con un Ipad. Carino, l'ho paragonato al Google Home mini che già possiedo, e sinceramente preferisco quest'ultimo. Comunque prodotto valido e divertente, al prezzo di lancio è sicuramente da provare.' There are three arrows pointing from labels above to specific parts of the review: 'CONFIGURAZIONE' points to 'configurarlo', 'PREZZO' points to 'prezzo scontato', and 'ESTETICA' points to 'Carino'. Below the review text, it says '7 persone l'hanno trovato utile'.

fred02

★★★★☆ Carino, da provare.

Recensito in Italia il 2 novembre 2018

Style: Echo Dot | Colore: Antracite | Configurazione: Solo dispositivo | **Acquisto verificato**

Ingoloso dal prezzo scontato ho acquistato Alexa e mi è puntualmente arrivato il 30 ottobre. Ho avuto difficoltà a configurarlo perché l'app. Alexa, sul mio OP5 non permetteva la configurazione. Questo è un problema che sembra affliggere anche i dispositivi Xiaomi. L'ho poi configurato con un Ipad. Carino, l'ho paragonato al Google Home mini che già possiedo, e sinceramente preferisco quest'ultimo. Comunque prodotto valido e divertente, al prezzo di lancio è sicuramente da provare.

7 persone l'hanno trovato utile



4) ASPECT-BASED SENTIMENT ANALYSIS

- 2 fasi principali:
 - 1) estrazione delle caratteristiche
 - 2) **classificazione** del sentiment per ciascuna caratteristica

CONFIGURAZIONE=NEG

PREZZO=POS

ESTETICA=POS

The image shows a product review for the Echo Dot. The reviewer is 'fred02', who gave it a 4-star rating and wrote 'Carino, da provare.' The review is dated November 2, 2018, and is marked as 'Acquisto verificato'. The product specifications listed are 'Style: Echo Dot', 'Colore: Antracite', and 'Configurazione: Solo dispositivo'. The review text describes a difficulty with configuring the device on an Android phone, which was resolved using an iPad. The reviewer concludes that the device is 'Carino' (cute) and 'valido e divertente' (valid and fun), and that it was worth the launch price. Annotations with arrows link specific parts of the text to sentiment labels: 'CONFIGURAZIONE=NEG' points to the configuration difficulty, 'PREZZO=POS' points to the launch price, and 'ESTETICA=POS' points to the word 'Carino'.

fred02

★★★★☆ Carino, da provare.

Recensito in Italia il 2 novembre 2018

Style: Echo Dot | Colore: Antracite | Configurazione: Solo dispositivo | **Acquisto verificato**

Ingolosito dal prezzo scontato ho acquistato Alexa e mi è puntualmente arrivato il 30 ottobre. Ho avuto difficoltà a configurarlo perché l'app. Alexa, sul mio OP5 non permetteva la configurazione. Questo è un problema che sembra affliggere anche i dispositivi Xiaomi. L'ho poi configurato con un Ipad. Carino, l'ho paragonato al Google Home mini che già possiedo, e sinceramente preferisco quest'ultimo. Comunque prodotto valido e divertente, al prezzo di lancio è sicuramente da provare.

7 persone l'hanno trovato utile



5) STANCE DETECTION

- Individuare il punto di vista generale (*posizione*) di una persona verso una cosa, un'idea, un evento
 - a favore o contro*
 - applicazioni in ambito politico: e.g., referendum 2016




A FAVORE

CONTRO



6) IRONY DETECTION

- Individuare contenuti ironici
 - l'ironia trasmette l'opposto del significato letterale
 - invertitore di polarità
 - satira, sarcasmo, parodia: è possibile distinguerli automaticamente?



Non riesco a trovare nani e ballerine nel governo Monti.
Ci deve essere un errore!! :)

[Translate Tweet](#)



7:53 AM · Nov 17, 2011 · [Tweetbot for iOS](#)



anagrammando Mario Monti no [#rimontiamo](#). Spread
480, la borsa rifiata e il mio mal di gola migliora.

[Translate Tweet](#)

11:59 AM · Nov 11, 2011 · [Twitter Web Client](#)



Mario Monti? non era il nome di un antipasto?
[#FullMonti](#) [#laresadeiconti](#) [#elezioni](#) #308

[Translate Tweet](#)

10:55 PM · Nov 9, 2011 · [Twitter Web Client](#)



SARCASMO

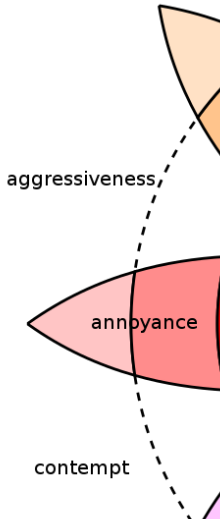
E vai adesso con Mario Monti/Superman, crisi finita,
stipendi in aumento, e riforme. Grazie Stato! Se non ci
fossi bisognerebbe inventarti!

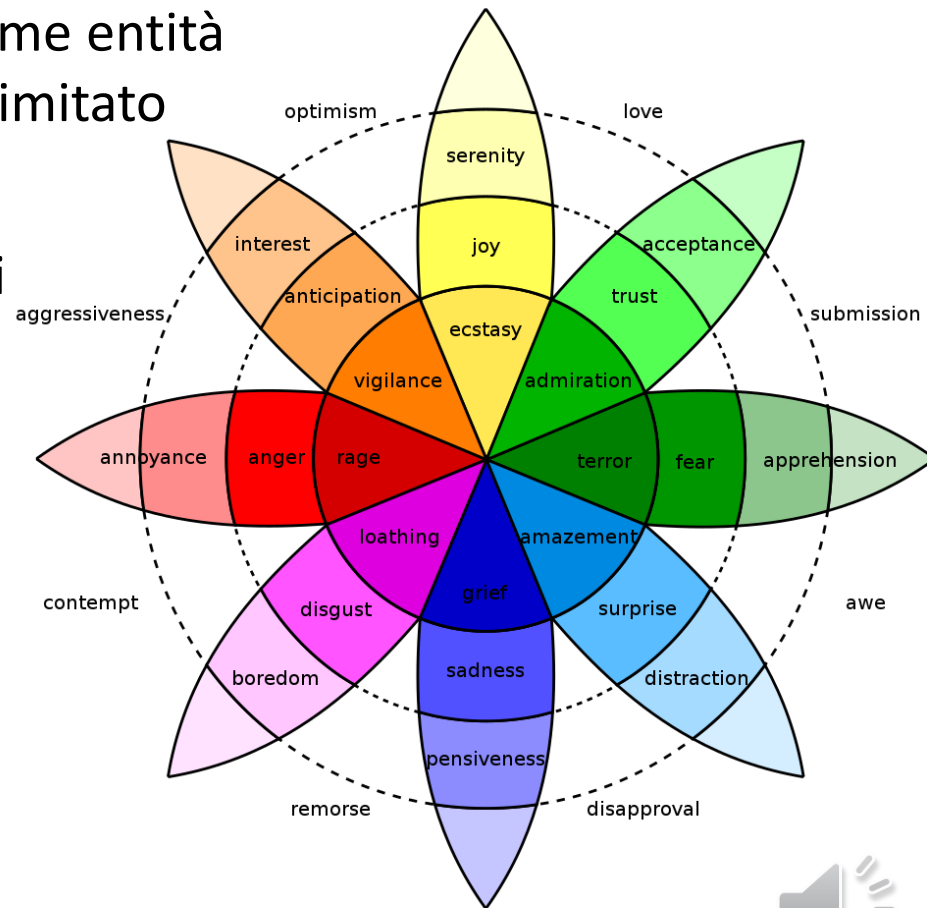
[Translate Tweet](#)

10:16 PM · Nov 12, 2011 · [Twitter Web Client](#)



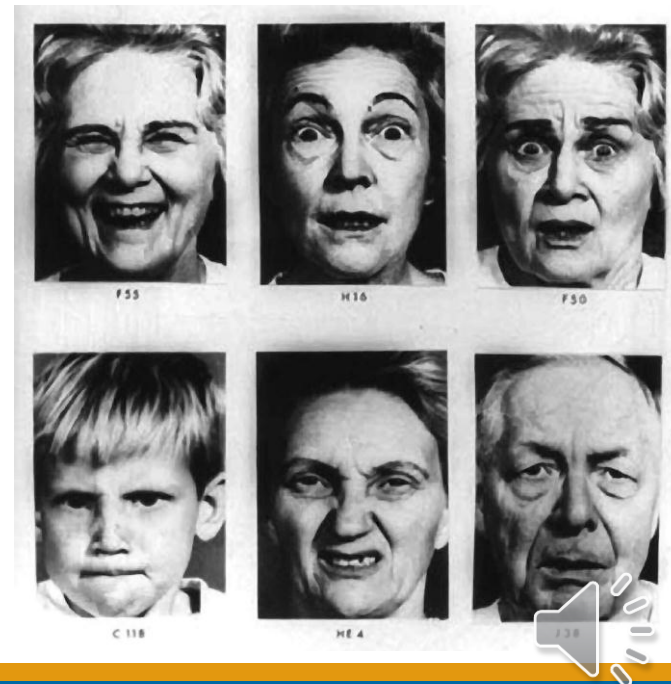
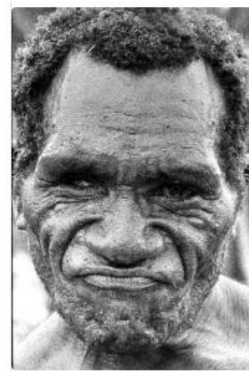
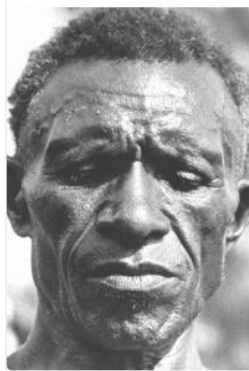
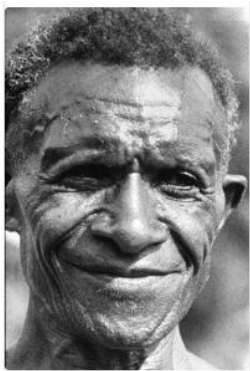
7) EMOTION ANALYSIS

- 2 approcci
 - 1) identificare le emozioni viste come entità innate ed universali di numero limitato
 - la ruota di Plutchik:
4 coppie di emozioni primarie che si combinano
 - *gioia* contro *dolore*
 - *rabbia* contro *paura*
 - *fiducia* contro *disgusto*
 - *sorpresa* contro *attesa*
- 
- A partial view of Plutchik's Wheel of Emotions. It shows three segments of the wheel: an orange segment at the top labeled 'aggressiveness', a red segment in the middle labeled 'annoyance', and a pink segment at the bottom labeled 'contempt'. Dashed lines indicate the boundaries between these segments and the rest of the wheel.



7) EMOTION ANALYSIS

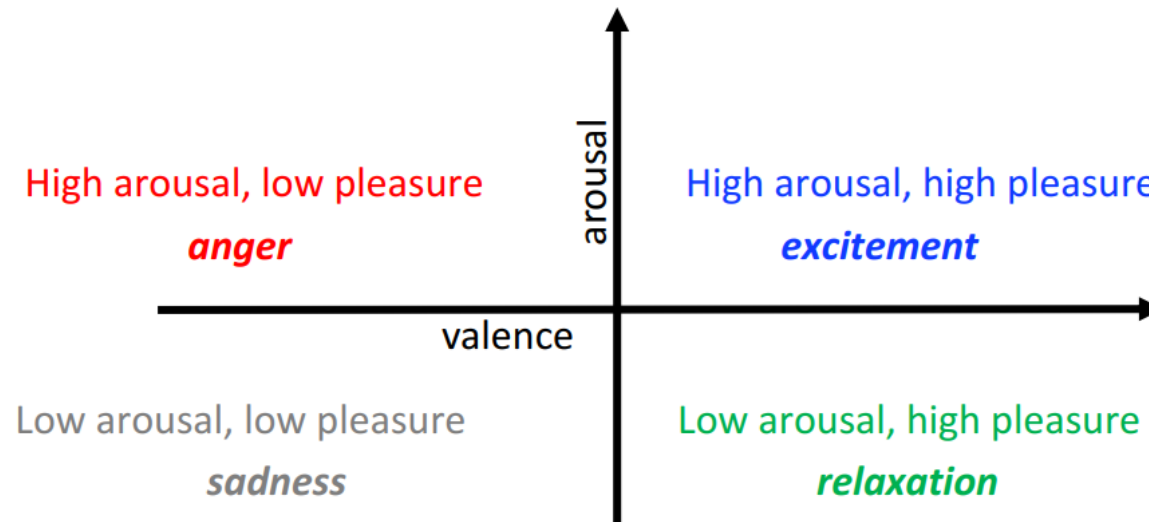
- 2 approcci
 - 1) identificare le emozioni viste come entità innate ed universali di numero limitato
 - gli studi di Ekman sull'universalità delle emozioni: rabbia, disgusto, tristezza, gioia, paura, sorpresa



<https://www.paulekman.com/>

7) EMOTION ANALYSIS

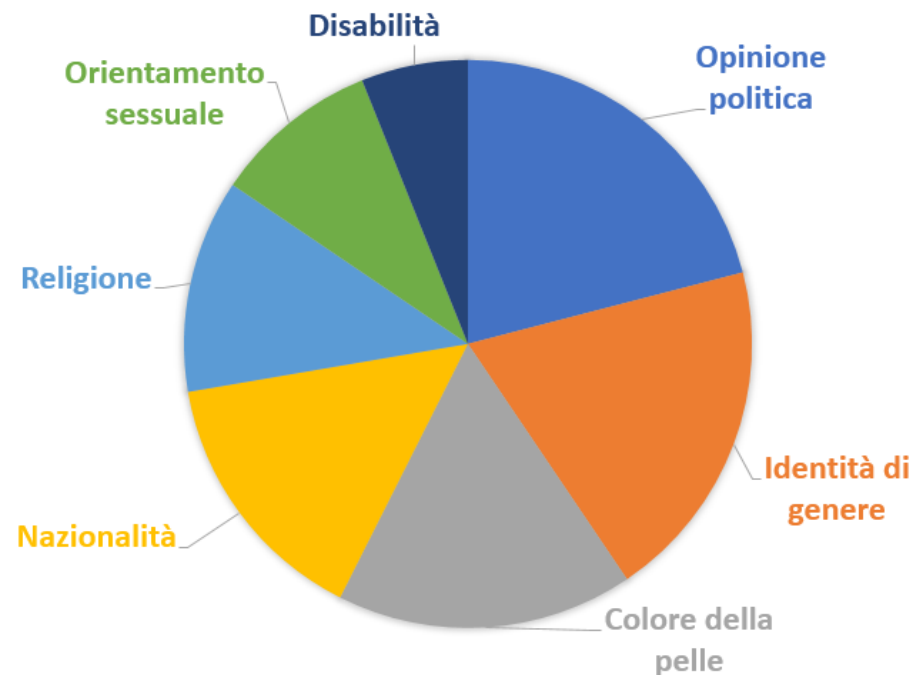
- 2 approcci
 - 1) identificare le emozioni viste come un numero limitato di unità innate ed universali
 - 2) identificare varie dimensioni delle emozioni
 - Valence: piacevolezza dell'emozione
 - Arousal: intensità dell'emozione
 - Dominance: grado di controllo esercitato dall'emozione



8) HATE SPEECH DETECTION

- Hate Speech: “*Qualsiasi **forma d’espressione** che fomenti l’**odio** contro una persona o un gruppo di persone a causa di caratteristiche che le accomunano.*” (Mendel, 2012 «Does international law provide for consistent rules on hate speech? »)

eMORE, Monitoring and Reporting Online Hate Speech in Europe, 2017



8) HATE SPEECH DETECTION



Cercasi urgentemente questa cretina per dirle in faccia quanto sia cretina! Grazie [#stopislam](#) [#STOPINVASIONE](#) [#stopidioti](#)

[Translate Tweet](#)

- Islamofobia

- Cyberbullismo



la [#raggi](#) sì che è una patata bollente altro che quella figa di legno della [#boschi](#) o quella cassapanca della [#boldrini](#) 😂😂
[#Libero](#) @vfeltri

10/02/17, 11:16

- Misoginia

