

Introduzione alla Sentiment Analysis

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, also called opinion mining, is the field of study that analyzes people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes.»

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

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 eccezionali 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

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



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

- **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**»
 - «la malattia può avere un decorso **acuto**»
 - «famoso per il suo spirito **acuto** e intelligente»

IL LINGUAGGIO DELLE OPINIONI

- Il linguaggio che usiamo per esprimere le nostre valutazioni soggettive è estremamente complesso, formato da molte componenti:
 - stesse parole, sensi diversi (**Word Sense Disambiguation**) -
Sentix lexicon

<i>PAROLAPoS</i>		<i>#senso</i>	<i>POS_score</i>	<i>NEG_score</i>	<i>polarity</i>	<i>intensity</i>
<i>acuto</i>	<i>a</i>	<i>00044760</i>	<i>0</i>	<i>0.5</i>	<i>-1.0</i>	<i>0.5</i>
<i>acuto</i>	<i>a</i>	<i>00650900</i>	<i>0.625</i>	<i>0.125</i>	<i>0.75</i>	<i>0.64</i>
<i>acuto</i>	<i>a</i>	<i>00803275</i>	<i>0.5</i>	<i>0.25</i>	<i>0.41</i>	<i>0.56</i>
<i>acuto</i>	<i>a</i>	<i>01744515</i>	<i>0.125</i>	<i>0</i>	<i>1.0</i>	<i>0.125</i>

IL LINGUAGGIO DELLE OPINIONI

- Il linguaggio che usiamo per esprimere le nostre valutazioni soggettive è estremamente complesso, formato da molte componenti:
 - stesse parole, sensi diversi, sentiment diversi (**Word Sense Disambiguation**) - Sentix lexicon

<i>PAROLA</i>	<i>PoS</i>	<i>#senso</i>	<i>POS_score</i>	<i>NEG_score</i>	<i>polarity</i>	<i>intensity</i>
<i>acuto</i>	<i>a</i>	00044760	0	0.5	-1.0	0.5
<i>acuto</i>	<i>a</i>	00650900	0.625	0.125	0.75	0.64
<i>acuto</i>	<i>a</i>	00803275	0.5	0.25	0.41	0.56
<i>acuto</i>	<i>a</i>	01744515	0.125	0	1.0	0.125

di malattia, che insorge con una certa violenza e ha un decorso rapido ed intenso

di facoltà intellettuale o di una sua manifestazione, che sa comprendere e valutare con prontezza le situazioni

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»*

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



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. 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 »



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...

The image shows a screenshot of a movie review interface. At the top left, it says 'no_data'. Below that is the title 'l'arte di essere americani' followed by four orange stars. The review text is in Italian. A green arrow points from the first sentence to the word 'POSITIVE' in green capital letters. A red arrow points from the last sentence to the word 'NEGATIVE' in red capital letters. The review text is: '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. [+]

At the bottom right, there is a poll section with two buttons: 'Sì' and 'No', both showing '0%' and the text 'd'accordo?' below them. The date 'martedì 29 luglio 2014' is in the top right corner.

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

Sì 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: '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.' There are three arrows pointing to specific parts of the review: one from 'CONFIGURAZIONE' pointing to 'configurarlo', one from 'PREZZO' pointing to 'prezzo scontato', and one from 'ESTETICA' pointing to '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

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 good and worth the price. Annotations with arrows point to specific parts of the review: a red arrow points to 'CONFIGURAZIONE=NEG' (pointing to the configuration difficulty), a green arrow points to 'PREZZO=POS' (pointing to the price), and another green arrow points to 'ESTETICA=POS' (pointing to the reviewer's positive opinion on the device's appearance).

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?



7) EMOTION ANALYSIS

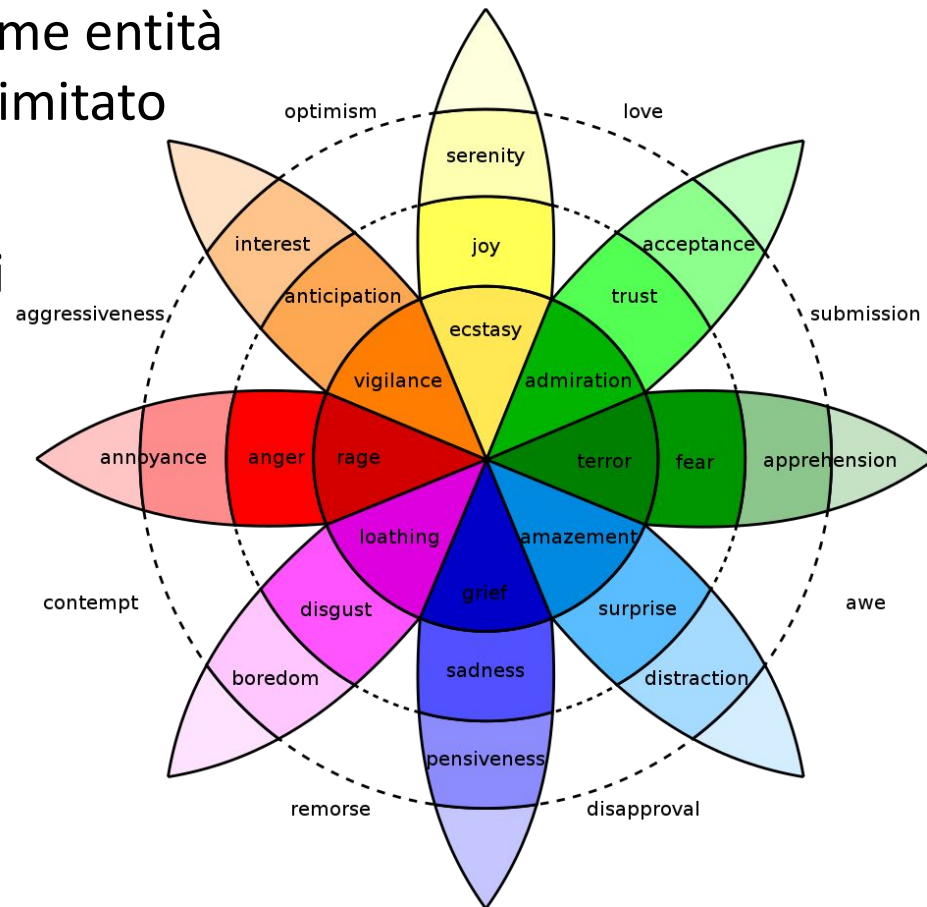
- 2 approcci

1) identificare le emozioni viste come entità innate ed universali di numero limitato

- la ruota di Plutchik:

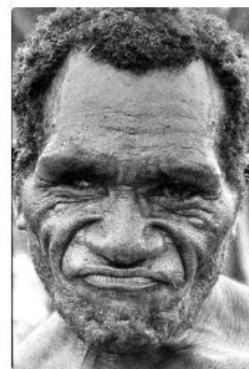
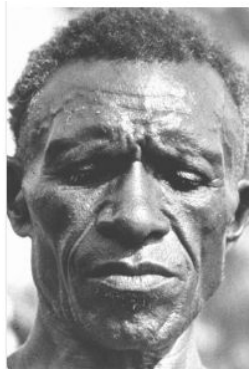
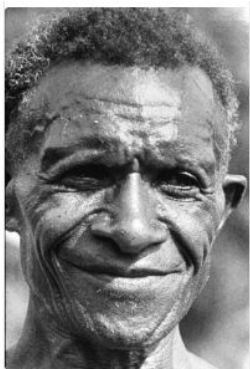
4 coppie di emozioni primarie che si combinano

- *gioia* contro *tristezza*
- *rabbia* contro *paura*
- *fiducia* contro *disgusto*
- *sorpresa* contro *trepidazione*



7) EMOTION ANALYSIS

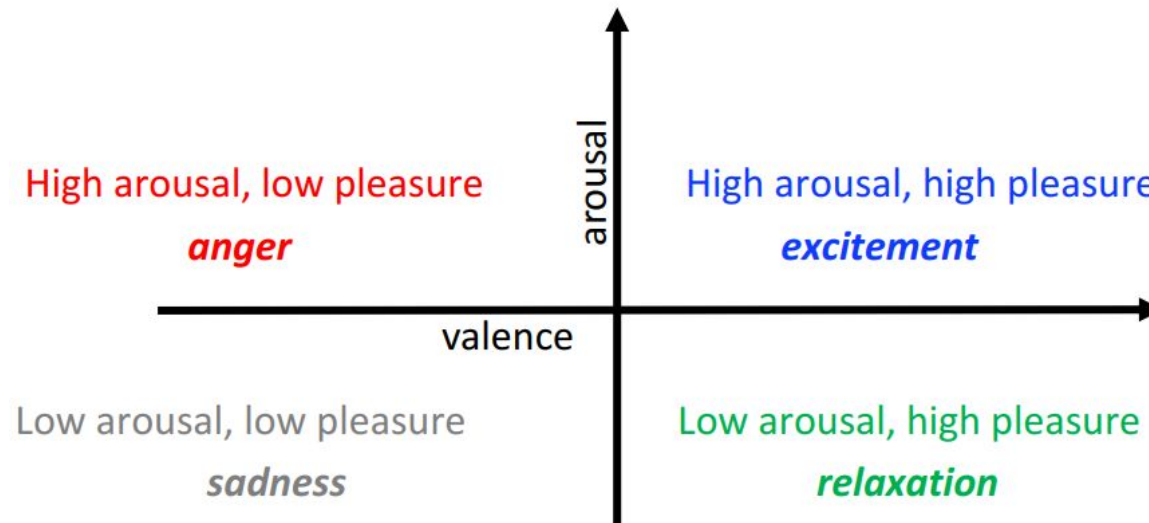
- 2 approcci
 - 1) identificare le emozioni viste come entità innate ed universali di numero limitato
 - gli studi di Ekman sull'universalità delle emozioni: rabbia, disgusto, tristezza, gioia, paura, sorpresa



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

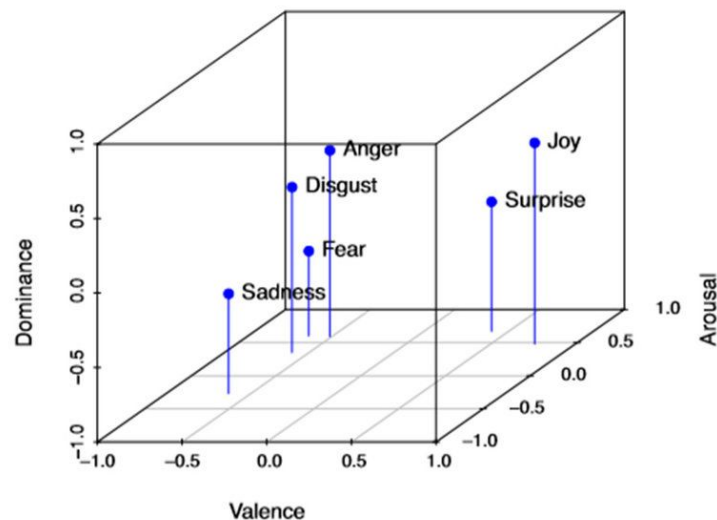
7) EMOTION ANALYSIS

- 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



7) EMOTION ANALYSIS

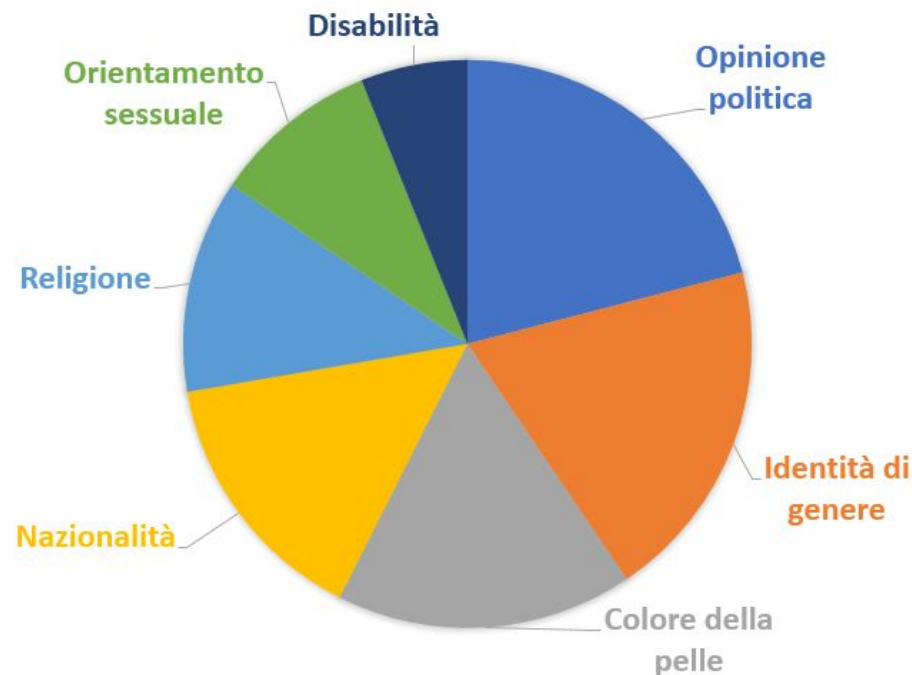
- 2 approcci
 - 1) identificare le emozioni viste come un numero limitato di unità innate ed universali
 - 2) identificare varie dimensioni delle emozioni
 - Valence: piacevolezza dell'emozione
 - Arousal: intensità dell'emozione
 - Dominance: grado di controllo esercitato dall'emozione



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



- Islamofobia

- Cyberbullismo

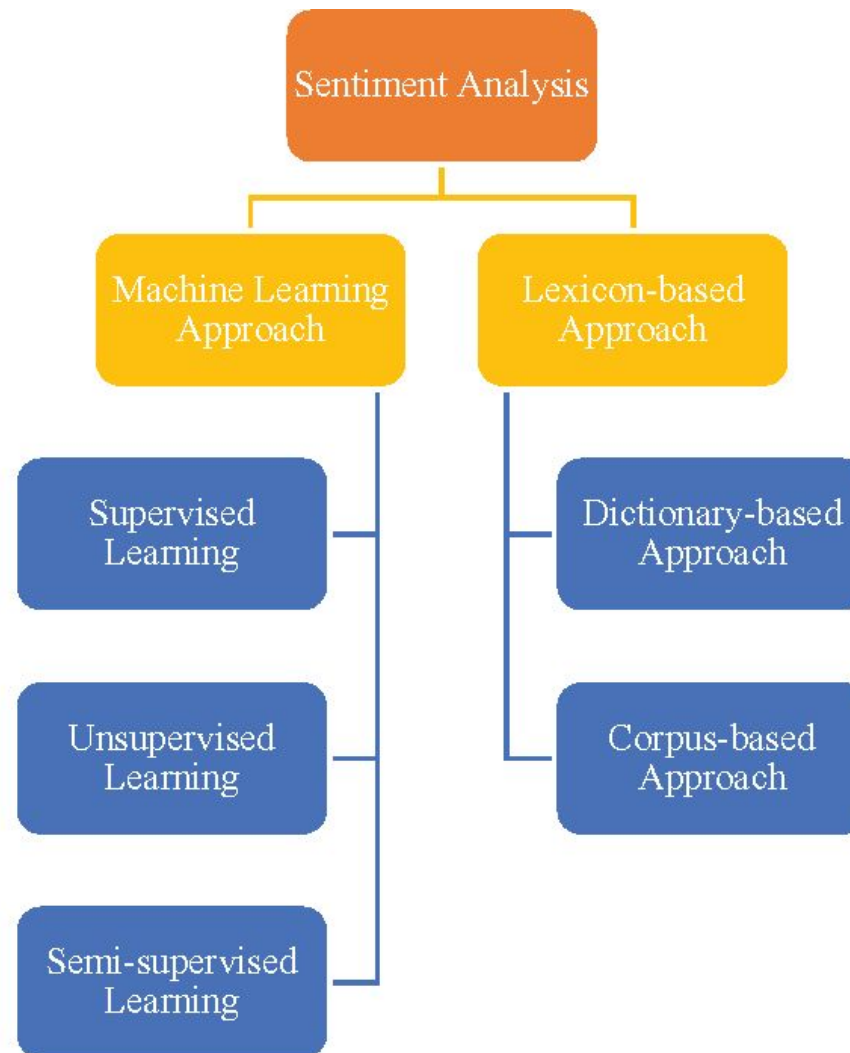


la #raggi sì che è una patata bollente altro che quella figa di legno della #boschi o quella cassapanca della #boldrini 😂😂
#Libero @vfeltri

- Misoginia

10/02/17, 11:16

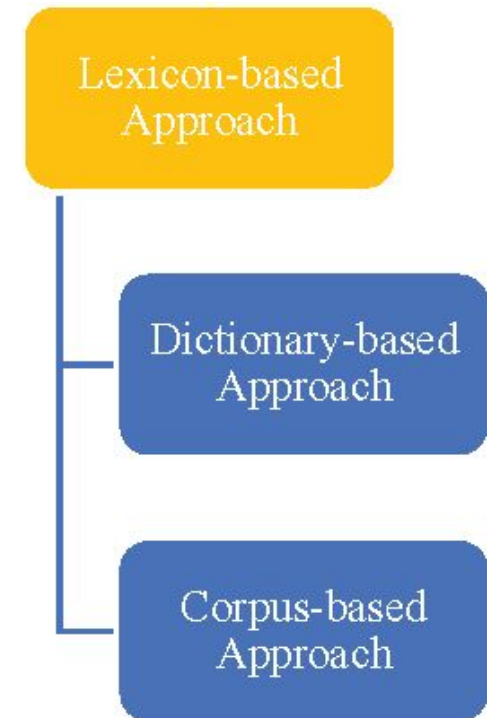
COME SI RISOLVE UN TASK DI SENTIMENT ANALYSIS



APPROCCI BASATI SUL LESSICO

- Basati sull'intuizione che la polarità di testo può essere ottenuta sulla base della polarità delle parole che lo compongono
- Si basano su liste di token, espressioni o lemmi («*sentiment/polarity lexicon*»): ad ogni token o lemma sono associate delle categorie o dei valori numerici che ne quantificano la polarità
Esempi:

- *negative, neutral, positive*
- *very negative, negative, neutral, positive, very positive*
- *-1, 0, +1*
- *-2, -1, 0, +1, +2*
- *valori decimali tra 0 e 1*



COME SI CREANO LESSICI DI SENTIMENT

- **DICTIONARY-BASED:** si usano una lista di termini con associato il sentiment e dei dizionari in formato digitale in cui si cercano sinonimi e contrari di quei termini

- *paura* = **NEG**

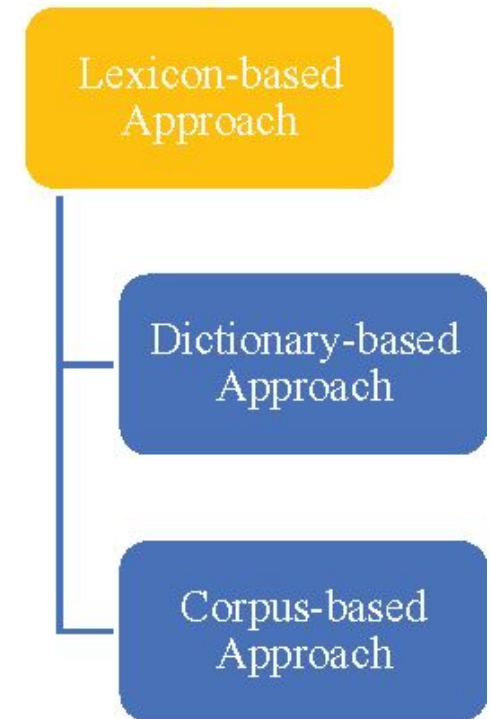
SINONIMI = *spavento, terrore* = **NEG**

CONTRARI = *coraggio, audacia* = **POS**

- **CORPUS-BASED:** si usano una lista di termini con associato il sentiment e delle collezioni di testi. Altri termini vengono aggiunti alla lista cercando contesti comuni o strutture sintattiche particolari

La macchina è bella E spaziosa

La macchina è bella MA difficile da guidare



LESSICI DI SENTIMENT PER L'ITALIANO - 1

- **OPENER:** più di 25.000 lemmi ed espressioni con PoS, 877 con sentiment assegnato manualmente, gli altri in maniera automatica


```
<LexicalEntry id="id_24468" partOfSpeech="noun">  
  <Lemma writtenForm="amico"/>  
  <Sense>  
    <Confidence score="0.5" method="automatic"/>  
    <Sentiment polarity="positive"/>  
    <Domain/>  
  </Sense>  
</LexicalEntry>
```

<https://github.com/opener-project/VU-sentiment-lexicon/tree/master/VUSentimentLexicon/IT-lexicon>

LESSICI DI SENTIMENT PER L'ITALIANO - 2

- **Sentix (Sentiment Italian Lexicon)**: più di 74.000 token ed espressioni con PoS, sentiment assegnato tramite allineamento di risorse pre-esistenti
 - sentiment diverso in base al significato

PAROLA	PoS	senso	POS_score	NEG_score	polarity	intensity
amico	n	09785042	0.25	0	1.0	0.25
amico	n	09877951	0.125	0	1.0	0.125
amico	n	10112591	0.125	0	1.0	0.125
amico	n	10686073	0.375	0	1.0	0.375



“someone who shares your feelings or opinions and hopes that you will be successful”

<http://valeriobasile.github.io/twita/sentix.gz>

LESSICI DI SENTIMENT PER L'ITALIANO - 3

- **Distributional Polarity Lexicon:** > 75.000 lemmi con PoS, sentiment associato automaticamente a partire da una collezione di tweet

parola::PoS positività, negatività, neutralità

amico::s 0.44531634,0.24274117,0.31194243

amico::a 0.7146159,0.15039273,0.1349914

#sanremo2012::h 0.26471627,0.28195098,0.45333272

#buonavita::h 0.71994877,0.052290898,0.22776033

<http://sag.art.uniroma2.it/demo-software/distributional-polarity-lexicon/>

LESSICI DI SENTIMENT PER L'ITALIANO - 4

- **SenticNet**: più di 23.000 lemmi ed espressioni rappresentanti concetti di senso comune con associati vari valori di polarità ed emozione

```
▼<rdf:RDF xmlns:rdf="http://w3.org/1999/02/22-rdf-syntax-ns#">
  ▼<rdf:Description rdf:about="http://sentic.net/api/it/concept/amico">
    <rdf:type rdf:resource="http://sentic.net/api/concept"/>
    <text xmlns="http://sentic.net">amico</text>
    ▼<semantics xmlns="http://sentic.net">
      <concept xmlns="http://sentic.net" rdf:resource="http://sentic.net/api/it/concept/amicizia"/>
      <concept xmlns="http://sentic.net" rdf:resource="http://sentic.net/api/it/concept/grande"/>
      <concept xmlns="http://sentic.net" rdf:resource="http://sentic.net/api/it/concept/convivente"/>
      <concept xmlns="http://sentic.net" rdf:resource="http://sentic.net/api/it/concept/si_basano_su_altra_persona"/>
      <concept xmlns="http://sentic.net" rdf:resource="http://sentic.net/api/it/concept/divertirsi_insieme"/>
    </semantics>
    ▼<sentic xmlns="http://sentic.net">
      <pleasantness xmlns="http://sentic.net" rdf:datatype="http://w3.org/2001/XMLSchema#float">0.72</pleasantness>
      <attention xmlns="http://sentic.net" rdf:datatype="http://w3.org/2001/XMLSchema#float">0</attention>
      <sensitivity xmlns="http://sentic.net" rdf:datatype="http://w3.org/2001/XMLSchema#float">0</sensitivity>
      <aptitude xmlns="http://sentic.net" rdf:datatype="http://w3.org/2001/XMLSchema#float">0.879</aptitude>
    </sentic>
    ▼<moodtags xmlns="http://sentic.net">
      <concept xmlns="http://sentic.net" rdf:resource="http://sentic.net/api/it/concept/felicità"/>
      <concept xmlns="http://sentic.net" rdf:resource="http://sentic.net/api/it/concept/ammirazione"/>
    </moodtags>
    ▼<polarity xmlns="http://sentic.net">
      <value xmlns="http://sentic.net">positive</value>
      <intensity xmlns="http://sentic.net" rdf:datatype="http://w3.org/2001/XMLSchema#float">0.533</intensity>
    </polarity>
  </rdf:Description>
</rdf:RDF>
```

<https://sentic.net/>

LESSICI DI SENTIMENT PER L'ITALIANO - 5

- **NRC-Emotion-Lexicon:** quasi 14.000 lemmi ed espressioni tradotti automaticamente dall'inglese, assegnazione di positività o negatività + assegnazione delle emozioni di Plutchik basata sull'annotazione manuale di vari annotatori non esperti

English	Italian	Pos	Neg	Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust
friend	amico	1	0	0	0	0	0	1	0	0	1
nausea	nausea	0	1	0	0	1	0	0	0	0	0
unnatural	innaturale	0	1	0	0	1	1	0	0	0	0

<https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm>

LESSICI DI SENTIMENT PER L'ITALIANO - 6

- **NRC-VAD-Lexicon:** più di 19.000 token ed espressioni tradotti automaticamente dall'inglese, assegnazione dei valori di Valence, Arousal e Dominance basata sull'annotazione manuale di vari annotatori non esperti

Word	Italian-it	Valence	Arousal	Dominance
friend	amico	0.906	0.413	0.573
pal	amico	0.812	0.426	0.596
nausea	nausea	0.104	0.781	0.273
natural	naturale	0.854	0.118	0.481
unnatural	innaturale	0.153	0.587	0.434

<https://saifmohammad.com/WebPages/nrc-vad.html>

LESSICI DI SENTIMENT PER L'ITALIANO - 7

- **Depeche Mood:** più di 82.000 parole ed espressioni con o senza PoS, assegnazione automatica di 5 emozioni derivata dai commenti online agli articoli del Corriere della Sera

TOKEN#PoS	INDIGNATO	PREOCCUPATO	TRISTE	DIVERTITO	SODDISFATTO	freq
raccoglimento#n	0.078	0.005	0.829	0.049	0.036	4
sims#n	0.0	0.0	0.0	0.687	0.312	15
fiabesco#a	0.012	0.0	0.028	0.0	0.959	3
squilibri#n	0.105	0.894	0.0	0.0	0.0	1
euroburocrati#n	0.972	0.0	0.0	0.0	0.027	3

<https://github.com/marcoguerini/DepecheMood/releases/tag/v2.0>

DATI ANNOTATI PER L'ITALIANO - 1

- **Sentipolc:** Sentiment Polarity Classification
 - Corpus di tweet annotati su quattro dimensioni:
 1. Soggettività: soggettivo versus oggettivo
 2. Polarità: negativo, positivo, mixed
 3. Ironia: ironico, non ironico
 4. Presenza di linguaggio figurato: polarità letterale, polarità non letterale

subj	opos	oneg	iro	lpos	lneg	description and explanatory tweet in Italian
0	0	0	0	0	0	objective <i>l'articolo di Roberto Ciccarelli dal manifesto di oggi</i> http://fb.me/1BQVy5Wak
1	0	0	0	0	0	subjective with neutral polarity and no irony <i>Primo passaggio alla #strabrollo ma secondo me non era un iscritto</i>
1	1	0	0	1	0	subjective with positive polarity and no irony <i>splendida foto di Fabrizio, pluri cliccata nei siti internazionali di Photo Natura</i> http://t.co/GWoZqbxAuS
1	0	1	0	0	1	subjective with negative polarity and no irony <i>Monti, ripensaci: l'inutile Torino-Lione inguaia l'Italia: Tav, appello a Mario Monti da Mercalli, Cicconi, Pont...</i> http://t.co/3CazKS7Y
1	1	1	0	1	1	subjective with both positive and negative polarity (mixed polarity) and no irony <i>Dati negativi da Confindustria che spera nel nuovo governo Monti. Castiglione: "Avanti con le riforme"</i> http://t.co/kIKnbFY7

<http://www.di.unito.it/~tutreeb/sentipolc-evalita16/data.html>

DATI ANNOTATI PER L'ITALIANO - 2

- **ABSITA:** Aspect-based Sentiment Analysis
 - Corpus di frasi tratte da recensioni di hotel di Booking annotati considerando 8 caratteristiche (e.g. pulizia, location) su due dimensioni:
 1. Presenza o meno della caratteristica
 2. Sentiment positivo o negativo per ciascuna caratteristica

cleanliness_ presence	cleanliness_ positive	cleanliness_ negative	location_ presence	location_ positive	location_ negative	sentence
0	0	0	1	1	0	La posizione, davvero invidiabile
0	0	0	1	0	1	Un po' fuori mano
1	1	0	1	1	0	Albergo pulito, in un'ottima posizione
1	0	1	0	0	0	Sporco a livelli colossali

<http://sag.art.uniroma2.it/absita/data/>

DATI ANNOTATI PER L'ITALIANO - 3

- **italian-sentiment-analysis**: Aspect-based Sentiment Analysis
 - Corpus di recensioni di smartphone

schermo[+],touch[+],volume[+],prezzo[+],ram[-]##Come si fa ad creare un cellulare simile : schermo bellissimo , touch bellissimo , volume bellissimo , prezzo incredibile , 2 alloggi per SIM Ma con così poca RAM Solo 512Mega

batteria[+],design[+],schermo[+]##Batteria impressionante, bel design e schermo piacevole.

<https://github.com/nicolaCirillo/italian-sentiment-analysis>

DATI ANNOTATI PER L'ITALIANO - 4

- **HaSpeeDe:** Hate Speech Detection
 - Corpus di tweet e post di Facebook annotati per la presenza di discorsi d'odio

Fonte	Tweet/Post	HateSpeech
Facebook	ITALIANI. BRUCIAMO LA STREGA	1
Facebook	Io mi auguro che gli italiani aprano finalmente gli occhi. @magdicristiano @angelo_ra_ Ci vuole la guerra per salvare	0
Twitter	l'Italia dai criminali filo islamici. https://t.co/TqHWKr74jB I traffici di organi seguono le rotte dei migranti. Summit @CasinaPioIV in Vaticano con medici giuristi politici da...	1
Twitter	https://t.co/kBGuhoHtpL	0

<http://www.di.unito.it/~tutreeb/haspeede-evalita18/data.html>

DATI ANNOTATI PER L'ITALIANO - 5

- **Italian Twitter Corpus of Hate Speech:** Hate Speech Detection
 - Corpus di tweet annotati su 7 dimensioni:
 1. Presenza di hate speech
 2. Target del tweet
 3. Grado di aggressività
 4. Grado di offensività
 5. Presenza di ironia
 6. Presenza di stereotipi
 7. Grado di intensità



Replying to @1282Ugo and @sputnik_italia

Bene,ricomincia la distruzione dei Porci terroristi

[Translate Tweet](#)

2:37 PM · Apr 12, 2017 · [Twitter for Android](#)

<https://github.com/msang/hate-speech-corpus>

DATI ANNOTATI PER L'ITALIANO - 6

- **Happy Parents:** Sentiment Polarity Classification
 - Corpus di tweet sul tema della fecondità e genitorialità annotati su 3 dimensioni:
 1. Soggettività
 2. Polarità con o senza ironia
 3. Argomento

Tweet	Annotazione	Argomento
@criharry bimba mia ti voglio tanto bene	POS	Being parent
Nella mia famiglia non parliamo tanto, però alla fine ci vogliamo bene!	MIXED	Daily live
Il trionfo di Renzi indebolisce vita e famiglia http://t.co/yAd1l10owN	NEG	Fertility & politics
Non ero preparato a fare il figlio, figuriamoci il genitore.	NEG	Becoming parents
Figliolo, lo vedi tutto questo disagio? Un giorno sarà tuo.	HUMNEG	Being parent

<https://github.com/mirkolai/Happy-Parents>

DATI ANNOTATI PER L'ITALIANO - 7

- **WhatsApp Dataset:** Cyberbullying Detection
 - Corpus di chat di WhatsApp raccolte tramite simulazione in classi seconde di scuole medie annotate in base a due dimensioni:
 1. Ruolo di chi scrive: e.g. vittima, bullo...
 2. Tipo di cyberbullismo: e.g. minacce, body shame, sessismo

Che fallito, vatti a chiudere nella fogna, finocchio!

[che fallito,]_{Insult: General Insult} [vatti a chiudere nella fogna,]_{Curse or Exclusion}
[finocchio!]_{Insult-Discrimination: Sexism}

<https://github.com/dhfbk/WhatsApp-Dataset>

QUANTO BENE FUNZIONANO I SISTEMI?

SISTEMI MACHINE LEARNING SULL'ITALIANO

- **Subjectivity Detection:** 0,74 F-score (2016)
- **Sentiment Classification** (Twitter): 0,66 F-score (2016)
- **Aspect-based SA - Category:** 0,81 F-score (2018) - 0,68 (2020)*
- **Aspect-based SA - Polarity:** 0,76 F-score (2018) - 0,62 (2020)*
- **Irony Detection:** 0,73 F-score (2018)
- **Irony Classification:** 0,52 F-score (2018)
- **Misogyny Detection:** 0,84 Accuracy (2018) - 0,74 F1 (2020)*
- **Misogyny Classification:** 0,41 Accuracy (2018)
- **Hate Speech Detection** (Twitter): 0,80 F-score (2018) - 0,81 (2020)
- **Hate Speech Detection** (Facebook): 0,83 F-score (2018)
- **Hate Speech Detection** (news): 0,77 F-score (2020)
- **Stance detection** (tweet): 0,67 F-score (2020)

Performance dei migliori sistemi di EVALITA 2016, 2018, 2020

APPROFONDIMENTI

Libro (pre-print):

<https://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf>

Lexicon-Based Methods for Sentiment Analysis:

https://www.mitpressjournals.org/doi/pdfplus/10.1162/COLI_a_00049

A Survey on Sentiment and Emotion Analysis for Computational Literary Studies: <https://arxiv.org/ftp/arxiv/papers/1808/1808.03137.pdf>

Overview of the EVALITA 2018 Aspect-based Sentiment Analysis task:

<http://ceur-ws.org/Vol-2263/paper003.pdf>

Overview of the EVALITA 2018 Task on Irony Detection in Italian Tweets:

<http://ceur-ws.org/Vol-2263/paper005.pdf>

Overview of the Evalita 2016 SENTIment POLarity Classification Task:

http://ceur-ws.org/Vol-1749/paper_026.pdf



GRAZIE!

Email: rachele.sprugnoli@unicatt.it

Twitter: [@RSprugnoli](https://twitter.com/RSprugnoli)

