crea
Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria

Sentiment Analysis: una panoramica

Cristina Bosco
Dipartimento di Informatica
Università degli Studi di Torino

UNIVERSITÀ DI TORINO

di.unito.it
DIPARTIMENTO DI INFORMATICA

CCC group al Dipartimento di Informatica di UniTo

- Il gruppo include vari membri del Dipartimento di Informatica con interessi di ricerca transdisciplinari
- I nostri progetti si focalizzano sullo sviluppo di risorse linguistiche annotate e strumenti per Sentiment Analysis, task di classificazione del testo e analisi morfo-sintattica
- Privilegiamo l'applicazione delle tecnologie del linguaggio a tematiche ad impatto sociale, lavorando ad esempio su hate speech detection e stereotype detection, e su tematiche relative all'ambiente

- Membri attivi della Associazione Italiana di Linguistica Computazionale
- Local chair di Language Resources and Evaluation Conference 2024, la conferenza internazionale della comunità che lavora sulle risorse linguistiche
- Il sottogruppo che lavora sulle tematiche ambientali include attualmente:

Valerio Basile

Cristina Bosco

Muhammad Okky Ibrohim

Finanziato da FSE REACT-EU per progetti di dottorato dedicati a GREEN topics

Overview

La Sentiment Analysis consiste in una procedura automatica di analisi del linguaggio naturale con l'obiettivo di classificare testi in termini di polarità (positiva, negativa o neutrale)

Le metodologie utilizzate nella Sentiment Analysis comprendono l'uso di lessici affettivi e di modelli computazionali (deep learning)

La Sentiment Analysis viene attualmente applicata anche per valutare le opinioni espresse sui social media nei confronti delle tematiche ambientali

La Sentiment Analysis consiste in una procedura automatica di a gio naturale con l'obiettivo di cl ermini di polarità (positiva, negativa o neutrale)

In cosa consiste la Sentiment Analysis?

Le metodologie utilizzate nella Sentiment Analysis comprendono l'uso di lessici affettivi e di modelli computazionali (deep learning)

La Sentiment Analysis viene attualmente applicata anche per valutare le opinioni espresse sui social media nei confronti delle tematiche ambientali

Overview

La Sentiment Analysis consiste in una procedura automatica di a gio naturale con l'obiettivo di cl ermini di polarità (positiva, negativa o neutrale)

In cosa consiste la Sentiment Analysis?

Le metodolog timent Analysis com ici affettivi e di modelli co arning)

Quali metodologie usa la Sentiment Analysis?

La Sentiment Analysis viene attualmente applicata anche per valutare le opinioni espresse sui social media nei confronti delle tematiche ambientali

Overview

La Sentiment Analysis consiste in una procedura automatica di a gio naturale con l'obiettivo di cl ermini di polarità (positiva, negativa o neutrale)

In cosa consiste la Sentiment Analysis?

Le metodolog timent Analysis com ici affettivi e di modelli co arning)

Quali metodologie usa la Sentiment Analysis?

La Sentiment mente applicata anc nioni espresse sui social me tematiche ambientali

In quali settori e perché si applica la Sentiment Analysis?

Sentiment Analysis e altre tecnologie

La Sentiment Analysis è soltanto una delle **tecnologie del linguaggio** che fanno ormai parte integrante dell'esperienza quotidiana di tutti, risultato di un'area di ricerca nata circa 70 anni fa, la linguistica computazionale (o natural language processing)

Sentiment Analysis e altre tecnologie

Filtraggio automatico dello spam

Traduzione automatica

Risposta automatica a domande

Sistemi di dialogo

Analisi morfologica e sintattica del testo

…

Sono tutte tecnologie del linguaggio, forme diverse di trattamento automatico del linguaggio umano, come la Sentiment Analysis!

Sentiment Analysis

Sentiment Analysis (or Opinion Mining) is "the computational study of opinions, sentiments and emotions expressed in text" (Bing Liu, 2012)

La SA appartiene alla famiglia dei task di classificazione automatica del testo, che consistono nell' attribuzione di una categoria (scelta in un dato insieme di categorie) ad ogni documento frase

La SA è il campo del trattamento automatico del linguaggio che si occupa di classificare i testi (ad esempio recensioni o tweet) in base al loro sentiment, ovvero se i testi stanno esprimendo una polarità positiva, negativa o neutra.

La più semplice forma di SA consiste nella rilevazione della polarità, cioè nell'identificazione di ogni testo come positivo o negativo nel suo complesso

Cosa c'è di meglio di un gol nel giorno del proprio compleanno?
Due gol nel giorno del proprio compleanno 😎🎂

Dato un testo in input il sistema restituisce il testo classificato come positivo o negativo.

Spesso viene considerata anche la possibilità che il testo sia espressione di polarità neutrale o mista.

Forme più raffinate di SA prevedono la classificazione del testo a grana più fine, consentendo l'identificazione di opinioni relative a specifici aspetti.
Si parla allora di Aspect-based SA o di Structured SA.

Il locale è bellissimo, luminoso e ben arredato. Ma non mi è piaciuto nessuno dei piatti che ho assaggiato. Non penso che ci tornerò.

location

food

Sentiment Analysis: come funziona

Gli strumenti di SA possono ottenere conoscenza utile a svolgere il task di classificazione della polarità da risorse esterne, in particolare da **lessici affettivi**

In un lessico affettivo ogni parola è associata alla sua polarità, espressa tramite un'etichetta o un valore scalare

Il locale è bellissimo, luminoso e ben arredato. Ma non mi è piaciuto nessuno dei piatti che ho assaggiato. Non penso che ci tornerò.

Sentiment Analysis: come funziona

Gli strumenti di SA possono ottenere conoscenza utile a svolgere il task di classificazione della polarità da risorse esterne, in particolare da lessici affettivi

In un lessico affettivo ogni parola è associata alla sua polarità, espressa tramite un'etichetta o un valore scalare

Il locale è **bellissimo**, **luminoso** e **ben arredato**. Ma **non** mi è **piaciuto nessuno** dei piatti che ho assaggiato. Non penso che ci tornerò.

Il riconoscimento della polarità di un intero testo dipende quindi dalla presenza di espressioni polarizzate e dalla loro composizione

Sentiment Analysis: come funziona

Gli strumenti di SA risultano ben più efficaci se oltre ad utilizzare informazioni lessicali apprendono conoscenza grazie ad algoritmi di **machine learning e deep learning.**

Durante la fase di addestramento (*training*) generano un **modello** del fenomeno da osservare, cioè della polarità, sulla base di tutte le informazioni disponibili in un campione di dati, detto *training set*

Se i testi usati nel *training* sono stati precedentemente annotati, rendendone esplicita la polarità, l'apprendimento risulta ancora più efficace

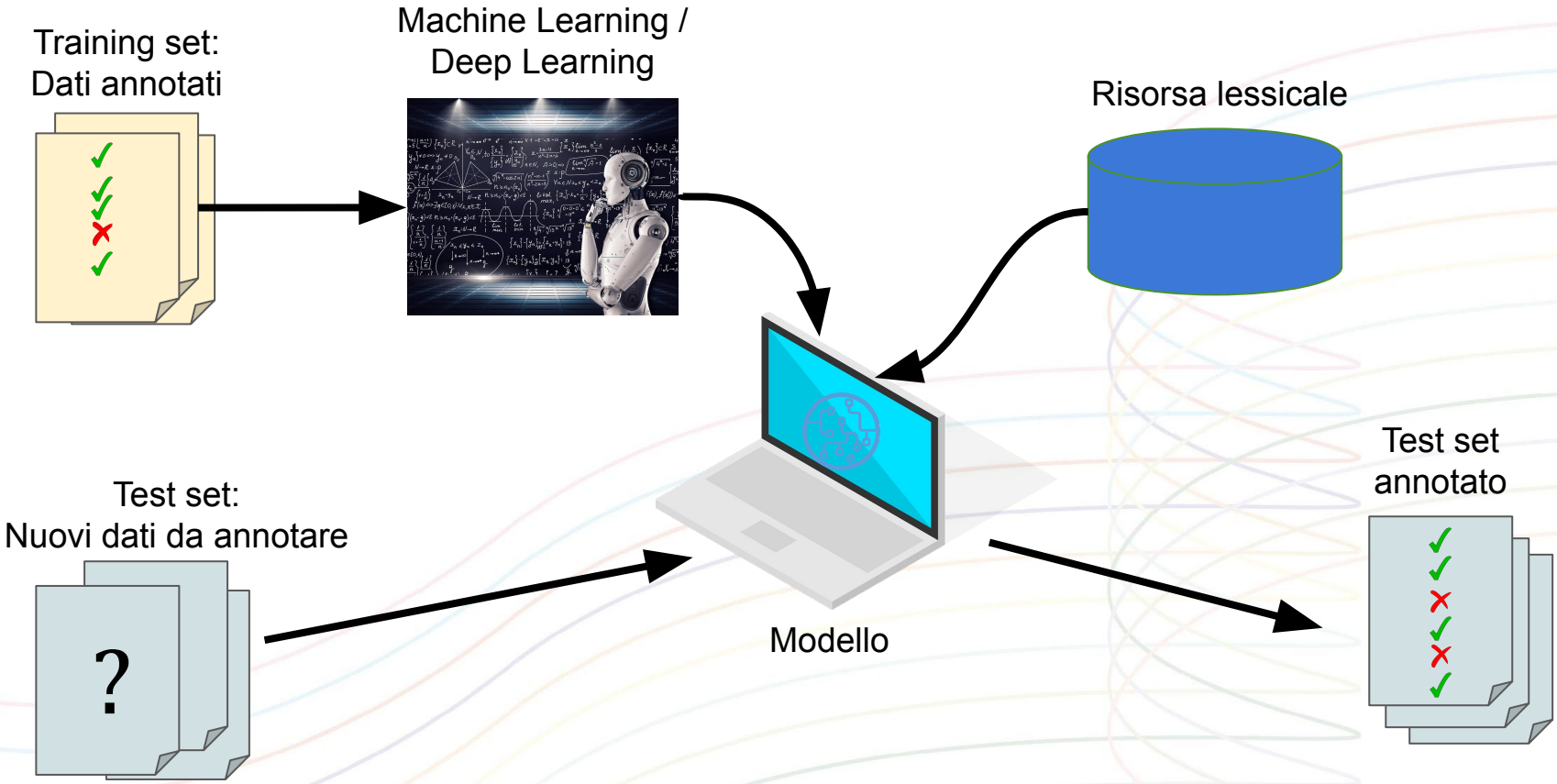
Sentiment Analysis: come funziona

Una volta costruito il modello, esso viene testato per valutarne la capacità di riconoscere correttamente la polarità in nuovi testi

La **valutazione** di uno strumento di SA avviene tramite il confronto con le prestazioni raggiunte da un essere umano nello svolgimento dello stesso task

Si prende quindi un nuovo campione di testi, lo si fa annotare dal sistema e in parallelo da un giudice umano, e poi si confrontano i risultati

Sentiment Analysis: come funziona



Sentiment Analysis: come funziona

L'allestimento di un sistema di SA comporta:

Raccolta di un dataset di testi

Creazione / reperimento del dizionario affettivo

Annotazione del training set e del test set

Applicazione dell'algoritmo di learning e creazione del modello

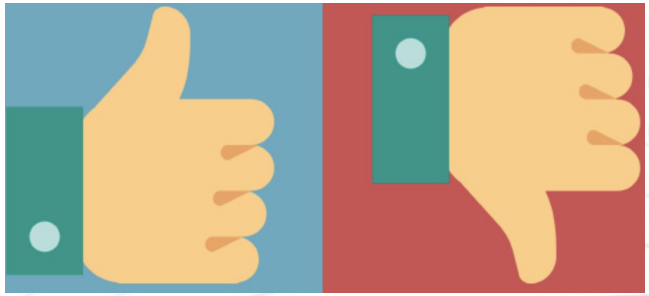
Valutazione del modello

Sentiment Analysis e linguaggio

I modelli del linguaggio creati nel contesto della linguistica computazionale e utilizzati nella SA riescono a tenere conto di una grande varietà di aspetti linguistici e della peculiare complessità del linguaggio

Il linguaggio è un sistema aperto in continua evoluzione e cambiamento in relazione ai diversi domini del discorso e generi testuali, uno specchio della società

La SA è stata applicata dal suo esordio per scopi di marketing, con l'obiettivo di carpire l'opinione degli utenti su specifici prodotti, interi brand o personaggi.



La SA consente di scoprire ed analizzare la *reputation* di aziende e politici per definire strategie di marketing e campagne elettorali più mirate e data-driven.

Sentiment Analysis: dove

Il principale campo di applicazione della SA è il web e le piattaforme di micro-blogging dove gli utenti delle varie community postano recensioni e commenti, cioè *user generated content*.

Tutti coloro che operano sui social media rilasciando commenti, opinioni, recensioni sono *opinion holder* in grado di modificare la percezione di una persona o di un brand.

Sentiment Analysis: in che settori e perché

Nell'ultimo decennio i task di classificazione del testo hanno assunto un ruolo sempre più importante nel contesto di **tematiche ad alto impatto sociale**, come la rilevazione dell'hate speech, della misoginia o della pedofilia.

La SA può giocare un ruolo anche nella sfida sociale in atto sulle **tematiche ambientali**?

Il problema:

I testi che affrontano tematiche ambientali sembrano essere caratterizzati da una relativa complessità in termini di varietà di argomenti trattati e sfaccettature nelle stesse

> La città ha una buona qualità dell'aria ma gli abitanti bruciano i rifiuti creando disagi per i soggetti allergici come me. Mi spiace vedere come le persone non apprezzino il valore dell'aria non inquinate … li dovremmo mandare tutti in una città inquinata per fargli provare cosa significa

Che risultati potrebbe produre la SA su questa tipologia di testi?

L'**analisi della leggibilità** (indice Gulpease) di un campione di testi pubblicati per informare i cittadini sulle tematiche green conferma che sono complessi e pienamente accessibili solo per lettori con un medio-alto livello di scolarità.

L'analisi è stata applicata ad un corpus che contiene:

- report della European Environment Agency (556 frasi in Italiano e 562 frasi in inglese)

- articoli sull'allevamento intensivo estratti da **CREAfuturo**, L'Informatore agrario e agricultura.it (725 frasi in italiano organizzate in 21 documenti)

Bosco, Ibrohim, Basile, Budi (2023 in press)
How green is Sentiment Analysis? Environmental Topics in Corpora at the University of Turin.
Proceedings of the ninth Italian Conference on Conputational Linguistics (Clic-It 2023)

L'applicazione della Sentiment Analysis alle tematiche ambientali è stata fino ad oggi molto limitata:

- sono stati testati quasi esclusivamente approcci basati su risorse lessicali e modelli pre-addestrati

- sono stati sviluppati pochi dataset annotati e solo per poche lingue

- tutti i dataset esistenti si limitano ad annotare la polarità a livello di interi documenti o frasi

Ibrohim, Bosco, Basile (2023 in press) *Sentiment Analysis for the Natural Environment: a Systematic Review*. ACM computing surveys

Il nostro progetto consiste nello sviluppare per l'italiano e l'inglese **un dataset annotato per la Structured Sentiment Analysis applicata alle tematiche ambientali**

Un approccio a grana più fine consente di annotare ed estrarre informazioni su chi esprime l'opinione, sul target a cui si rivolge l'opinione e quindi su cosa viene effettivamente criticato o apprezzato dai cittadini

La complessità della comunicazione ambientale dipende infatti anche dalla varietà degli enti coinvolti e delle tematiche su cui i cittadini si esprimono.

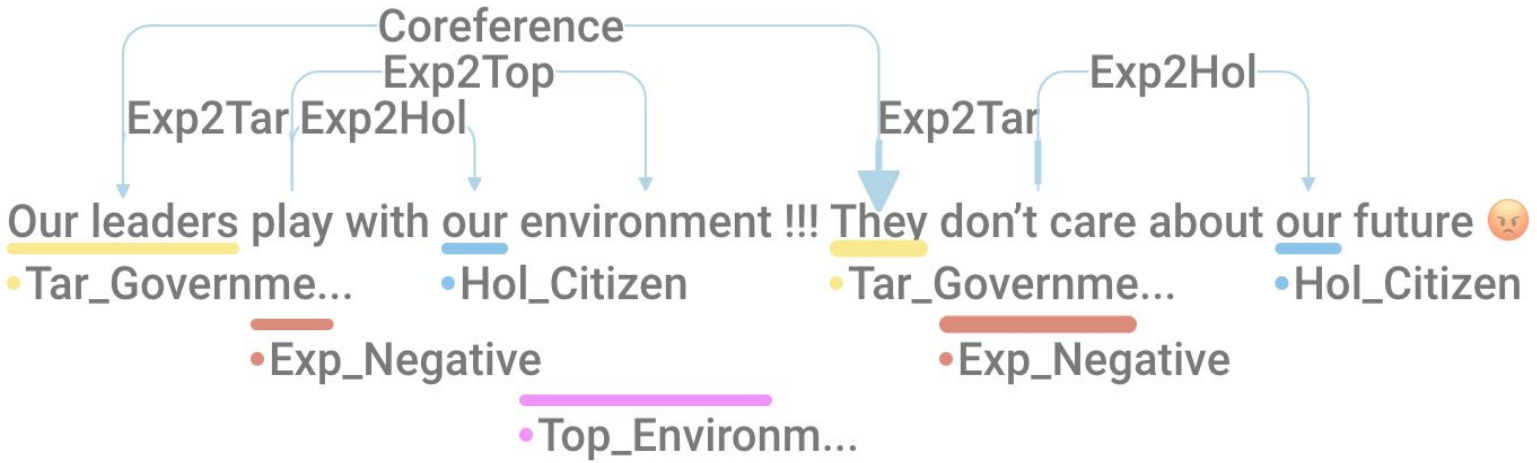
Il **dataset** contiene attualmente:

- ˜8.500 tweet in **italiano** raccolti tra febbraio e marzo 2022 contenti le keywords: Transizione energetica, Agenda 2030, Crisi climatica, Combustibili fossili, Deforestazione, Greenwashing, Riscaldamento globale, Impatto ambientale, Climate Change, Green Deal, sviluppo sostenibile, COP26, Energie rinnovabili

- ˜500.000 tweet in **inglese** raccolti a settembre 2022 con un set più ampio di keyword su 10 tematiche: Environment, Green, Sustainability, Food, Organism, Climate Change, Carbon, Energy,Waste, Pollution

I due dataset sono in fase di annotazione, quello inglese in crowdsourcing

crea
Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria



Sentiment Analysis e linguaggio

Il problema di come **comunicare** il cambiamento climatico diventa sempre più pressante. Il messaggio degli scienziati non ha fatto breccia nella società e l'urgenza di un'azione adeguata è percepita soltanto da una minoranza nella popolazione e nella classe dirigente

Di cosa abbiamo bisogno per fare un **salto di qualità nel livello della comunicazione**? Quali sono gli errori da evitare e quale linguaggio risulta più appropriato per toccare le corde delle persone?

La SA può aiutare a conoscere l'autentica opinione delle persone sulle tematiche ambientali?

Può collaborare con altre discipline per la costruzione di percorsi di comunicazione più efficaci sulle tematiche ambientali?

La SA può aiutare a scoprire i paradigmi negativi e i bias cognitivi presenti nel nostro modo di parlare dell'ambiente, che ostacolano la costruzione di una efficace comunicazione sulle tematiche ambientali

Conclusioni

La **psicologia cognitiva** lavora sulla comunicazione del cambiamento climatico facendo i conti con le caratteristiche del cervello umano e con i suoi limiti, la dissonanza cognitiva e i pregiudizi cognitivi (*cognitive biases*).

Conclusioni

L'**ecolinguistica** si propone di smontare i paradigmi negativi che tengono assieme le società per aprirsi a idee nuove.

Per consegnare un messaggio positivo nell'ambito del clima occorre ricorrere a metafore per ridurre la complessità, ma quali metafore? meglio paragonare l'effetto serra a una coperta o ad un forno?

Si deve operare una sorta di *reframing,* imperniando il discorso su valori ed emozioni, non sul linguaggio dell'economia e usando termini semplici ed immagini.

L'antropologia sottolinea che **non esiste una concezione della natura neutra ed universale**, ma che essa è dipendente dalla cultura e dalla storia della società di cui si fa parte.

«Il modo in cui l'Occidente moderno rappresenta la natura è la cosa meno condivisa nel mondo» (Marshall Sahlins)

La nostra visione dell'ambiente può cambiare?