Analisi del Sentiment

Lorenzo Scorzelli, Giovanni Borrelli

UNISA

[l.scorzelli1@studenti.unisa.it](mailto:l.scorzelli1@studenti.unisa.it), [g.borrelli17@studenti.unisa.it](mailto:g.borrelli17@studenti.unisa.it)

Obiettivi

L’obiettivo di questo progetto è di usare un modello di sentiment già addestrato per estrarre informazioni utili per lo studio delle intercettazioni criminali. Il progetto si basa su un dataset contenente delle intercettazioni criminali in lingua italiana, che verranno sottoposte al modello di sentiment per rilevare le emozioni presenti nel testo delle intercettazioni. Questo può essere utile a capire lo status attuale dei criminali intercettati, le loro motivazioni, le loro intenzioni e le loro relazioni. Per raggiungere questo obiettivo, il progetto prevede le seguenti fasi: la preparazione e la pulizia del dataset di intercettazioni criminali; l’applicazione del modello di sentiment al dataset di intercettazioni criminali; l’analisi e la visualizzazione dei risultati del modello di sentiment; la discussione delle implicazioni e delle limitazioni del modello di sentiment per il campo della legge.

Analisi del Sentiment

L'analisi del sentiment rappresenta una disciplina fondamentale nell'ambito dell'elaborazione del linguaggio naturale (NLP), mirata a decifrare e interpretare le emozioni e le opinioni espresse nei testi scritti. Questo campo, in continua evoluzione, ha attraversato diverse fasi, dalla semplice assegnazione di polarità basata su dizionari di parole chiave alle sofisticate tecniche di machine learning che caratterizzano l'attuale panorama della sentiment analysis.Inizialmente, le prime metodologie di analisi del sentiment si basavano su regole linguistiche e liste di parole predefinite, assegnando a ciascuna parola una polarità predefinita. Questi approcci, sebbene rudimentali, costituivano un primo tentativo di tradurre il linguaggio umano in un formato comprensibile per le macchine. Tuttavia, con l'avanzamento della capacità computazionale, il campo ha abbracciato con entusiasmo l'apprendimento automatico (machine learning), aprendo la strada a nuove prospettive.L'introduzione dei dizionari di sentimenti è stata una pietra miliare nell'evoluzione dell'analisi del sentiment. Modelli come il VADER Sentiment Analysis hanno implementato questa approccio, arricchendo i dizionari con informazioni sulla polarità e sull'intensità di ciascuna parola. Questa metodologia più sofisticata consente una valutazione più precisa, tenendo conto del contesto e delle sfumature linguistiche. Il VADER, in particolare, ha dimostrato di essere un pilastro nell'analisi del sentiment, ispirando numerosi sviluppi e applicazioni in vari settori.Le applicazioni pratiche dell'analisi del sentiment sono ampie e trasversali. Nel marketing, le imprese sfruttano questa tecnologia per comprendere il feedback dei clienti, adattare le strategie di vendita e valutare l'accoglienza dei propri prodotti e servizi. Nei social media, l'analisi del sentiment è diventata uno strumento essenziale per monitorare l'opinione pubblica su questioni di rilievo, identificare trend emergenti e valutare la reputazione di brand e personaggi pubblici.Tuttavia, nonostante i progressi significativi, l'analisi del sentiment si trova ancora di fronte a diverse sfide. La comprensione di espressioni idiomatiche, la gestione delle sfumature culturali e il riconoscimento del contesto rimangono aree di miglioramento. L'auspicio è che con l'impiego di modelli sempre più sofisticati e l'integrazione di approcci di machine learning avanzati, il campo possa superare queste sfide e offrire una comprensione ancora più profonda del linguaggio delle emozioni.In questa panoramica sull'analisi del sentiment, abbiamo delineato l'evoluzione di questa disciplina, dalle sue radici tradizionali fino alle attuali frontiere del machine learning. Nei prossimi paragrafi, ci immergeremo in due modelli specifici: il VADER Sentiment Analysis e il framework Spacy. Analizzeremo come questi modelli affrontano le sfide dell'analisi del sentiment, contribuendo a plasmare il modo in cui comprendiamo e interpretiamo il linguaggio delle emozioni nel contesto del NLP.

# VADER

Uno dei punti salienti nella storia dell'analisi del sentiment è l'avvento del VADER Sentiment Analysis, un modello concepito da Hutto e Gilbert nel 2014. Il nome VADER è un acronimo di Valence Aware Dictionary and sEntiment Reasoner, che sottolinea la sua duplice natura di dizionario di sentimenti e ragionatore di sentimenti. Questo modello ha rivoluzionato l'analisi del sentiment attraverso l'implementazione di un approccio basato su dizionari e regole sofisticate.Il cuore del VADER è il suo vasto dizionario di sentimenti, nel quale ogni parola è associata a un punteggio sulla scala della polarità. Questo dizionario, creato con attenzione per riflettere la complessità delle espressioni umane, è un elemento cruciale che consente al modello di valutare il sentiment complessivo di un testo. A differenza di molti altri approcci, il VADER non si limita a classificare le parole come positive o negative; tiene anche conto dell'intensità del sentimento, catturando le sfumature del linguaggio in modo più accurato.Le regole di punteggio del VADER giocano un ruolo fondamentale nel determinare il sentiment complessivo di una frase. Ad esempio, il modello è in grado di rilevare negazioni e altre costruzioni linguistiche che possono invertire il significato di una parola. Inoltre, le emoticon, spesso utilizzate nelle comunicazioni online, vengono integrate nel processo di analisi, aggiungendo ulteriori dettagli alla comprensione del sentiment complessivo.Le applicazioni pratiche del VADER Sentiment Analysis sono diversificate e ampiamente adottate in vari settori. Nel marketing, le aziende utilizzano il modello per valutare la reazione dei clienti ai loro prodotti e servizi. Nei social media, il VADER è impiegato per monitorare il sentiment pubblico su determinati argomenti, identificare tendenze emergenti e valutare la percezione di un brand o di un'organizzazione.L'impatto del VADER nella ricerca accademica e nell'industria è innegabile. Il modello ha ispirato lo sviluppo di nuovi approcci nell'analisi del sentiment, spingendo la comunità scientifica a esplorare ulteriormente le potenzialità dei modelli basati su dizionari e regole. A livello pratico, il VADER è diventato uno strumento essenziale per molte organizzazioni che cercano di comprendere il sentiment del cliente e ottimizzare le loro strategie di comunicazione.Nonostante il suo successo, il VADER Sentiment Analysis non è immune da sfide. La gestione di contesti culturali diversi e la comprensione di espressioni idiomatiche rappresentano ancora ambiti di miglioramento. Tuttavia, gli sviluppatori stanno continuamente affinando il modello, integrando approcci di machine learning avanzati per affrontare queste sfide e migliorare ulteriormente le sue prestazioni.In conclusione, il VADER Sentiment Analysis ha aperto nuove prospettive nell'analisi del sentiment, fornendo un modello flessibile e preciso per decodificare le emozioni nel linguaggio umano. La sua combinazione di dizionari di sentimenti e regole di punteggio ha contribuito significativamente a una comprensione più approfondita del modo in cui le parole trasmettono il sentiment, gettando le basi per ulteriori innovazioni nel campo dell'elaborazione del linguaggio naturale.

# Spacy

Spacy rappresenta una pietra miliare nell'ambito dell'analisi del linguaggio naturale (NLP), offrendo un framework potente e flessibile che si estende ben oltre l'analisi del sentiment. Lanciato nel 2015, Spacy è diventato un punto di riferimento per la comunità di ricerca e per le applicazioni industriali, grazie alla sua efficienza, versatilità e alla capacità di gestire complessità linguistiche in modo accurato.

Il cuore di Spacy è il suo motore di analisi del linguaggio naturale, che utilizza algoritmi di machine learning per scomporre il testo in token, frasi e concetti grammaticali. A differenza di molti altri framework, Spacy è progettato per essere veloce ed efficiente, rendendolo ideale per l'analisi di grandi volumi di testo in tempo reale. La sua architettura modulare permette agli sviluppatori di selezionare le componenti specifiche di cui hanno bisogno, rendendo Spacy altamente personalizzabile per soddisfare le esigenze specifiche di ogni progetto.

Una delle caratteristiche distintive di Spacy è la sua capacità di integrare modelli pre-addestrati per svolgere una varietà di attività linguistiche, tra cui l'analisi del sentiment. I modelli di analisi del sentiment in Spacy possono essere adattati e personalizzati per contesti specifici, consentendo agli sviluppatori di ottenere risultati altamente specializzati. Questa flessibilità è un vantaggio significativo, specialmente quando si affrontano problemi complessi di analisi del testo.

Oltre all'analisi del sentiment, Spacy offre funzionalità avanzate come il riconoscimento di entità (NER), l'analisi grammaticale, e l'estrazione di concetti chiave. Queste capacità estese rendono Spacy una scelta preferita per una vasta gamma di applicazioni, dall'elaborazione di documenti legali alla creazione di chatbot intelligenti. La sua architettura modulare e la presenza di modelli pre-addestrati semplificano notevolmente il processo di sviluppo, riducendo il tempo e la complessità necessari per implementare soluzioni NLP avanzate.

L'ampia adozione di Spacy sia in ambito accademico che industriale testimonia della sua efficacia. Aziende di diverse dimensioni e settori integrano Spacy nelle loro applicazioni per ottenere una comprensione più profonda del linguaggio umano e migliorare le interazioni con gli utenti. L'ecosistema di Spacy continua a crescere, con la comunità di sviluppatori che contribuisce a nuovi modelli e miglioramenti costanti.

Nonostante i suoi successi, Spacy affronta costantemente nuove sfide. La comprensione di contesti altamente specifici e la gestione di lingue meno diffuse sono ambiti in cui il framework continua a evolversi. Gli sviluppatori di Spacy rispondono prontamente alle sfide emergenti, lavorando per migliorare continuamente la precisione e l'adattabilità del framework.

In conclusione, Spacy si posiziona come un pilastro fondamentale nell'analisi del linguaggio naturale, offrendo un approccio flessibile e potente per esplorare le complessità del linguaggio umano. La sua capacità di adattarsi a una vasta gamma di contesti, combinata con la sua efficienza e versatilità, continua a renderlo una scelta di spicco per coloro che cercano di sfruttare appieno le potenzialità dell'NLP nelle loro applicazioni e progetti.

# Feel-It

Feel-It è uno strumento innovativo che analizza i sentimenti espressi sui social media in lingua italiana. È stato sviluppato da due ricercatori, Federico Bianchi e Debora Nozza, e un professore associato, Dirk Hovy, dell’Università Bocconi di Milano.

Questo strumento open source è in grado di riconoscere e analizzare i sentimenti e le emozioni trasmesse attraverso i social media in italiano. È stato progettato per essere facilmente installabile in qualsiasi pipeline e utilizzabile da chiunque.

Il funzionamento di Feel-It si basa sull’analisi di tweet in italiano. Gli sviluppatori hanno estratto le emozioni da alcuni tweet riguardanti vari argomenti di tendenza su Twitter. Hanno poi analizzato e classificato manualmente circa 2000 tweet associando a ciascuno un’emozione (gioia, rabbia, tristezza, paura). Questo dataset è stato utilizzato per addestrare un modello neurale chiamato “BERT” in italiano.

Feel-It è in grado di mappare le emozioni delle persone suddividendole in due cluster: positivo/negativo (“joy” per il positivo e le altre a negativo), riuscendo a replicare il modello BERT. I risultati hanno dimostrato che Feel-It è un modello che predice bene la sentiment analysis e le emozioni nel testo.

Un punto importante da sottolineare è che Feel-It non è limitato all’analisi dei soli testi dei social media. I risultati hanno dimostrato che questo strumento può essere utilizzato anche con testi non legati ai social media, ampliando così il suo campo di applicazione.

Feel-It può rivelarsi molto utile per aziende e PMI che hanno una presenza digitale, in quanto permette di capire il trend degli utenti e di analizzare come le persone rispondono a certi eventi sui social.

Il dataset e il modello di Feel-It sono disponibili in open source sul web e sono descritti in un paper accademico peer-reviewed che è stato presentato a WASSA 2021 durante l’undicesimo Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis della conferenza EACL.

Feel-It si dimostra quindi adatto all’obiettivo di questo progetto. Oltre a identificare se il sentimento espresso in un testo è positivo o negativo, Feel-It è in grado di distinguere tra diverse emozioni. Questa capacità fornisce informazioni più dettagliate che possono aiutare a comprendere lo stato emotivo del criminale analizzando la chiamata telefonica intercettata. Inoltre, l’analisi delle emozioni può contribuire a capire il tipo di legame che esiste tra i due interlocutori in chiamata.

# Differenze tra VADER,Spacy e Feel-It

L’analisi del linguaggio naturale (NLP) è stata rivoluzionata da diversi modelli, tra i quali: il VADER Sentiment Analysis, il framework Spacy e Feel-It. Mentre tutti e tre si concentrano sull’esplorazione delle complessità del linguaggio umano, le loro architetture, metodologie e applicazioni variano significativamente.

Il VADER Sentiment Analysis si distingue per il suo approccio basato su dizionari e regole sofisticate. Questo modello è principalmente adatto all’inglese e si concentra sulla valutazione del sentiment complessivo di un testo, prendendo in considerazione sia la polarità delle parole che l’intensità del sentimento.

D’altra parte, Spacy adotta un approccio più modulare e flessibile basato su machine learning. Il framework offre un motore di analisi del linguaggio naturale che utilizza algoritmi di machine learning per scomporre il testo in token, frasi e concetti grammaticali. Spacy offre un’ampia suite di funzionalità NLP, tra cui l’analisi grammaticale, il riconoscimento di entità e l’analisi del sentiment.

Feel-It, sviluppato da ricercatori dell’Università Bocconi di Milano, è uno strumento open source che analizza i sentimenti espressi sui social media in lingua italiana. Utilizza un modello neurale chiamato “BERT” addestrato su un dataset di tweet in italiano. Feel-It è in grado di mappare le emozioni delle persone suddividendole in due cluster: positivo/negativo (“joy” per il positivo e le altre a negativo). A differenza di VADER e Spacy, Feel-It fornisce informazioni sulle emozioni e non solo sulla positività o negatività del sentimento.

Per lo scopo del progetto, Feel-It risulta essere il più adatto. Questo perché, oltre a fornire informazioni sulla positività o negatività del sentimento, Feel-It è in grado di distinguere tra diverse emozioni. Inoltre, Feel-It è stato addestrato su frasi italiane, il che lo rende particolarmente utile per l’analisi del linguaggio naturale in italiano.

### [Intestazione 3].

[Includere un punto alla fine di un titolo normale. Si noti che è possibile includere paragrafi consecutivi con titoli specifici, se appropriato.]

#### [Intestazione 4].

[Quando si usano i titoli, non saltare i livelli. Se è necessario un titolo 3, 4 o 5 senza testo a seguire prima del titolo successivo, aggiungere un punto alla fine del titolo e quindi iniziare un nuovo paragrafo per il sottotitolo e il relativo testo.] (Cognome, Anno)

##### [Intestazione 5].

[Come tutte le sezioni del documento, i riferimenti iniziano in una pagina specifica. La pagina di riferimenti che segue è stata creata con la caratteristica Citazioni e bibliografia, disponibile nella scheda Riferimenti. Questa funzionalità include un'opzione di stile che formatta i riferimenti per APA sesta edizione. È anche possibile usarla per aggiungere citazioni nel testo collegate all'origine, come quelle mostrate alla fine di questo paragrafo e nel paragrafo precedente. Per personalizzare una citazione, fare clic con il pulsante destro del mouse e scegliere Modifica citazione.] (Cognome, Anno)

Riferimenti

Cognome, N. (Anno). Titolo articolo. *Titolo rivista*, Pagine da-a.

Cognome, N. (Anno). *Titolo libro.* Nome città: Nome editore.

Note a piè di pagina

1[Aggiungere le eventuali note a piè di pagina nella pagina appropriata dopo i riferimenti. Per i requisiti di formattazione APA, è facile digitare note e rimandi alle note a piè di pagina personalizzati. Per formattare un rimando a una nota a piè di pagina, selezionare il numero e quindi, nella raccolta Stili della scheda Home, fare clic su Rimando nota a piè di pagina. Per il corpo di una nota a piè di pagina, come in questo esempio, si usa lo stile del testo Normale. (Nota: se si elimina una nota a piè di pagina di esempio, ricordarsi di eliminare anche il rimando nel testo. Si trova alla fine del paragrafo con il Titolo 2 di esempio nella prima pagina del contenuto del corpo in questo modello.)]

Tabelle

Tabella 1

[Titolo tabella]

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Intestazione colonna | Intestazione colonna | Intestazione colonna | Intestazione colonna | Intestazione colonna |
| Intestazione riga | 123 | 123 | 123 | 123 |
| Intestazione riga | 456 | 456 | 456 | 456 |
| Intestazione riga | 789 | 789 | 789 | 789 |
| Intestazione riga | 123 | 123 | 123 | 123 |
| Intestazione riga | 456 | 456 | 456 | 456 |
| Intestazione riga | 789 | 789 | 789 | 789 |

Nota: [posizionare tutte le tabelle del documento nell'apposita sezione, dopo i riferimenti (e, se applicabile, le note a piè di pagina). Iniziare una nuova pagina per ogni tabella e includere un numero e un titolo per ognuna, come illustrato in questa pagina. Tutto il testo descrittivo visualizzato in una nota dopo la tabella, come in questo caso. Usare lo stile Tabella/Figura disponibile nella raccolta Stili della scheda Home per ottenere la spaziatura tra la tabella e la nota. Per le tabelle in formato APA si può usare l'interlinea singola o di 1,5 righe. Includere un'intestazione per ogni riga e colonna, anche se il contenuto sembra ovvio. Per questo modello è stato configurato uno stile di tabella predefinito che soddisfa le linee guida APA. Per inserire una tabella, fare clic su Tabella nella scheda Inserisci.]

Titolo figure:

Figura 1. [Includere tutte le figure in specifiche sezioni, dopo i riferimenti e, se applicabile, le note a piè di pagina e le tabelle. Includere una didascalia numerata per ogni figura. Usare lo stile Tabella/Figura per semplificare la spaziatura tra la figura e la didascalia.]

Per altre informazioni su tutti gli elementi della formattazione APA, consultare il manuale dello stile APA, sesta edizione.