

Opinion Mining

Carmine Romaniello

`carmine.romaniello@studenti.unimi.it`

Università degli studi di Milano, Via Celoria n.18, Milano, Italy

Abstract. In questo paper viene affrontato il problema di individuare le diverse opinioni espresse riguardo le singole entità all'interno di una recensione (**Opinion Mining**). Per ogni coppia (entità, opinione) individuata viene poi calcolata la **polarità** dell'opinione per determinarne il grado di giudizio e classificarla come positiva, negativa o neutra. Il dataset utilizzato viene fornito da **Yelp** ed è liberamente scaricabile in rete.

Keywords: Opinion mining · Sentiment analysis

1 Introduzione

Le opinioni ricoprono un ruolo fondamentale per noi umani perchè sono in grado di influenzare fortemente i nostri pensieri e i nostri comportamenti, infatti ogni qualvolta ci ritroviamo a dover prendere una decisione siamo spesso interessati a conoscere l'opinione altrui. Basti pensare ad un'impresa che vorrebbe sempre sapere cosa ne pensano i clienti riguardo i propri prodotti o servizi così come i singoli clienti vorrebbero sapere l'opinione degli altri prima di acquistare o usufruire di un servizio.

La crescita esponenziale dei social media sul Web ha permesso di ottenere una grande mole di dati grazie al quale la ricerca in questo campo è avanzata molto più rapidamente. Ad oggi esistono sistemi in grado di analizzare opinioni in qualunque dominio di applicazione, alcuni ad esempio riescono addirittura a determinare la percentuale di consenso politico o il futuro andamento di un titolo in borsa. Questo spiega perchè l'interesse verso questo campo è in continua crescita a tal punto da indurre alcune compagnie a svilupparne uno proprio.

Il Natural Language Processing (NLP) è un campo di studio e di ricerca il cui obbiettivo è quello di rendere le macchine abili nel comprendere il linguaggio naturale dell'uomo espresso in forma scritta e parlata. **Opinion Mining** (OM) è una branca di NLP che si occupa di individuare ed analizzare le opinioni e il pensiero che un individuo esprime. Molte volte OM viene confuso con **Sentiment Analysis** ed entrambi vengono utilizzati in modo interscambiabile, ma in realtà ognuno di essi riguarda un aspetto ben definito del problema. Come documentato in [1], con il termine opinion si intende l'espressione di un giudizio oggettivo o soggettivo verso un'entità mentre il sentiment è una reazione generata da una sensazione/emozione provata dall'individuo. Spesso la ricerca delle

opinioni viene focalizzata sulla parte soggettiva che invece si riferisce al sentimento provato nel momento in cui l'opinione viene espressa e questo provoca ambiguità.

Un campo di applicazione molto rilevante in cui vengono utilizzate le tecniche appena citate è quello delle recensioni. Ad oggi,, per veicolare meglio le nostre scelte, siamo molto interessati a leggere recensioni scritte da altri riguardo un prodotto, un esercizio commerciale, un luogo ecc. La stessa attività commerciale è molto interessata a sapere cosa scrivono i clienti nelle recensioni e che giudizio hanno verso un prodotto o un servizio offerto per poter migliorare. La lettura di ogni singola recensione richiederebbe però troppo tempo perciò ognuna di esse è accompagnata da un voto che ha lo scopo di fornire un giudizio complessivo e di sintesi (inserito dall'utente e solitamente espresso in numero di stelle da 1 a 5). I singoli voti vengono poi aggregati in modo tale che si possa sapere subito quanti voti positivi e negativi sono stati dati e farsi subito un'idea. Purtroppo in questo modo si ricava solamente una valutazione generale senza sapere quali sono le opinioni effettivamente espresse.

In questo paper viene documentato lo sviluppo di un sistema in cui si applica OM per analizzare le recensioni ed estrapolare le informazioni contenute all'interno. In ogni recensione siamo interessati a sapere cosa è stato espresso di positivo e/o di negativo, ma ancora di più a sapere se in una recensione molto positiva vi è qualcosa di negativo e viceversa (**outliers**). Il sistema prende in input una recensione da cui estrae le diverse opinioni riguardo le singole entità come coppie (entità, opinione). Per ogni coppia estratta ne determina la **polarità** (positiva, negativa o neutra) e successivamente aggrega le polarità delle singole opinioni per ottenere un giudizio complessivo sulla recensione paragonabile alle stelle.

L'obiettivo è quello di ottenere rapidamente le informazioni contenute come lista di opinioni ed evitare la lettura. La tecnica adoperata per l'estrazione delle opinioni (affrontata nella sezione 2) si basa sull'analisi sintattica delle frasi facendo uso di regole logiche e grammaticali. Per la determinazione della polarità vengono utilizzati dei tool preesistenti che in seguito vedremo in modo più approfondito.

1.1 Stato dell' arte

Come detto precedentemente vi è molto interesse in questo campo di ricerca infatti sono presenti molti lavori in letteratura in cui viene affrontato il problema (come documentato nel libro [1]). Tutti questi concordano sul modellare un'opinione come una quintupla $\langle e, a, p, h, t \rangle$ in cui **e** è un'entità, **a** è un aspetto dell'entità **e**, **p** è la polarità dell'opinione sull'aspetto **a**, **h** è il possessore dell'opinione e **t** è il momento temporale in cui l'opinione è stata espressa [1, 3]. I metodi usati per identificarla variano a seconda del dominio di applicazione, ma soprattutto in base al livello di granularità che si vuole raggiungere. In generale esistono 3 livelli qui elencati in ordine di granularità crescente:

- **Document Level:** Si assume che l'intero documento esprime opinioni riguardo una singola entità quindi si ottiene solamente una valutazione generale positiva, negativa o neutra sull'intero documento. Si presta bene quando si è

interessati ad ottenere una valutazione molto generale, superficiale e poco dettagliata.

- **Sentence Level:** Come il precedente, ma a livello di frase. Ricavando le valutazioni di ogni singola frase è possibile aggregarle per ottenere un giudizio complessivo sull'intero documento più preciso rispetto al livello precedente.
- **Entity-Aspect Level:** Nei livelli precedenti non è possibile sapere con esattezza cosa è stato espresso di positivo e non. In questo invece si arriva fino all'opinione stessa analizzando le relazioni tra le parole che compongono la frase per ottenere anche l'entità e l'aspetto a cui l'opinione si riferisce. E' quindi un approccio molto granulare che permette di ottenere più informazioni con maggior precisione, ma allo stesso tempo è anche molto più complesso.

Indipendentemente dal livello i principali indicatori utilizzati per determinare la polarità sono le cosiddette **opinion words** che insieme costituiscono l'**opinion lexicon**. Queste sono le parole comunemente usate per esprimere un'opinione (es. good, bad, great, horrible) e all'interno del lexicon vengono associate alla rispettiva polarità positiva o negativa. La necessità di un **opinion lexicon** è uno dei principali problemi dell'OM ancora non del tutto risolto, infatti in letteratura sono presenti molte proposte di approccio per la sua costruzione le cui principali sono di tipo: manuale (meno utilizzato perchè troppo time consuming), Dictionary-based e Corpus-Based.

2 Modello e sviluppo

In questo lavoro l'estrazione avviene al livello più granulare (Entity-Aspect Level) e si utilizza un **opinion lexicon Dictionary-Based** precostruito e già edichettato. Per lo scopo che si vuole raggiungere saremo interessati solamente alle tre componenti **e**, **a**, **p** della quintupla, pertanto si abuserà della definizione formale e si useranno i termini **aspetto** per indicare sia entità che aspetti e **opinione** per indicare l'insieme delle opinion words che determinano la **polarità**. Per ogni aspetto si estraggono diverse opinioni e la polarità finale associata all'aspetto dipenderà da quelle delle singole opinioni.

Le feature selezionate sono: le parti del discorso (POS), le dipendenze sintattiche (DEP), le Opinion Rules, le negazioni che invertono la polarità (Shifter) e gli avverbi che la alterano (Modifier). L'approccio seguito è basato sull'analisi testuale in cui vengono analizzate le singole POS e DEP tra loro esistenti. In generale, nonostante i diversi modi con cui è possibile costruire una frase, per le opinioni vengono usati gli aggettivi mentre per gli aspetti i sostantivi. Nell'opinione vengono considerati anche gli eventuali Modifier e Shifter che ne condizionano il senso. Per determinare univocamente le coppie aspetto-opinione si esaminano le DEP esistenti tra sostantivi e aggettivi e vengono ritenute valide solo se rispettano le Opinion Rule. Ottenuta la coppia si ricava la polarità dell'aggettivo che viene invertita se è presente uno Shifter e/o amplificata o ridotta a seconda del Modifier.

2.1 Descrizione del sistema

Il sistema è organizzato in una pipeline avente i seguenti stage:

- **Preprocessing:** suddiviso in due fasi. Nella prima vengono rimossi alcuni caratteri non alfanumerici come i simboli non di punteggiatura e la punteggiatura errata (es. il punto prima di "but" o di "and"). Nella seconda fase la recensione viene suddivisa in singole frasi che vengono elaborate singolarmente nel successivo stage della pipeline.
- **Aspect-Opinion Extraction:** è il core dell'intero sistema. Una coppia aspetto-opinione viene estratta se vengono rispettate alcune regole che impongono l'esistenza di determinate DEP. Tali regole sono descritte per semplicità nella tabella a pagina 8. L'opinione è composta da una tripla $\langle is_Shifted, Modifier, Adjective \rangle$ dove *is_Shifted* è un booleano, *Modifier* sono gli eventuali avverbi e *Adjective* è l'aggettivo.
- **Polarity Evaluation:** La polarità assume valori nell'intervallo $[-1,1]$. Per ogni coppia aspetto-opinione si determina la polarità dell'aggettivo che viene poi aumentata o diminuita a seconda del Modifier e invertita se *is_Shifted* è True. Considerando che un aspetto potrebbe avere più opinioni e in una recensione possono esserci più aspetti, la polarità finale della recensione viene calcolata come la media delle polarità di tutte le opinioni di tutti gli aspetti.
- **Review Classification:** La recensione viene classificata in negativa, positiva o neutra secondo la seguente metrica:
 - Negativa: polarità < -0.15
 - Neutra: $-0.15 < \text{polarità} < 0.15$
 - Positiva: polarità > 0.15
- **Outliers Detection:** Dalla lista di coppie aspetto-opinione vengono rilevate quelle incoerenti con il numero di stelle cioè quelle aventi una polarità negativa in una recensione positiva e viceversa.

Dettagli implementativi Il sistema è stato sviluppato in Python (ver. 3.7.2). Per la parte di estrazione è stata utilizzata la libreria `spacy` (ver. 2.3.0) con il modello per la lingua inglese "`en_core_web_lg`". La valutazione della polarità è stata implementata integrando il tool `VaderSentiment` con la libreria `wordnet` di `nltk` e l'opinion lexicon `SentiWordNet` [2]. L'integrazione è stata necessaria perchè dai test effettuati è emerso che spesso `VaderSentiment` assegna una polarità neutra ad aggettivi il cui senso può cambiare a seconda del contesto. Grazie a `wordnet` il senso viene disambiguato calcolando la polarità media di tutti i sensi che l'aggettivo può assumere.

3 Simulazione e test

Considerando la natura del problema di estrazione non è stato possibile effettuare un test automatizzato in quanto non esistono dataset adeguatamente etichettati con aspetti e opinioni. Le prestazioni sono state quindi valutate manualmente su un sottoinsieme limitato di recensioni con risultati soddisfacenti.

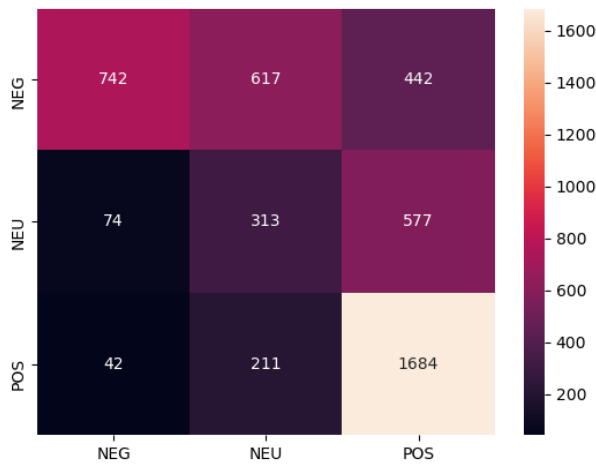
Per quanto riguarda invece la polarità e la conseguente classificazione è stato scelto di etichettare le recensioni in: Negativa (1 o 2 stelle), Neutra(3 stelle), Positiva (4 o 5 stelle) per confrontarle con il risultato predetto dal sistema.

3.1 Dataset

Il dataset utilizzato per gli esperimenti viene fornito liberamente dal famoso servizio di recensioni americano **Yelp** da cui ne sono state estrapolate casualmente un totale di 5000 (1000 per ogni stella). Il formato del dataset è di tipo **json** e da ogni recensione sono stati estrapolati solamente i campi utili per gli scopi di questo lavoro: **text** e **stars** . Si rimanda alla documentazione in [3] per i dettagli.

3.2 Risultati

I risultati sono stati valutati attraverso la matrice di confusione qui sotto riportata in cui sulle ordinate vi sono le classi etichettate mentre sulle ascisse quelle predette.



Dalla matrice vengono ricavate le seguenti metriche:

Classe	Support	Precision	Recall	F-measure
NEG	1801	0.86	0.41	0.56
NEU	964	0.27	0.32	0.30
POS	1937	0.62	0.87	0.73

Support totale: 4702

Accuracy : 58 %

Il support totale non corrisponde alle 5000 recensioni iniziali perchè quelle molto brevi che non contengono opinioni estraibili non vengono classificate.

I risultati ottenuti non sembrano essere abbastanza soddisfacenti come dimostra il valore di accuracy al di sotto del 60%. Il sistema tende a classificare troppo positivamente dando molti falsi positivi ed è molto impreciso con le recensioni neutre. Analizzando le recensioni scorrettamente classificate sono emerse le seguenti problematiche:

- I tools impiegati per il calcolo della polarità sono talvolta imprecisi a causa dell’ambiguità degli aggettivi. Ad esempio la parola ”slow” viene considerata neutra, ma riferita ad un servizio di ristorazione assume ovviamente un connotato negativo. Questo, purtroppo, è un problema aperto nel NLP che necessita di una trattazione a parte.
- A volte la polarità degli aggettivi non viene invertita in negativa perchè la negazione è implicita nella semantica della frase. Succede ad esempio quando l’utente usa un modo di dire oppure quando descrive un aspetto in senso negativo utilizzando aggettivi positivi o neutri.
- La recensione non è coerente con il numero di stelle.

Le prime due problematiche sono imputabili a carenze del sistema mentre l’ultima merita particolare attenzione. Le recensioni considerate incoerenti sono quelle che non rispecchiano il numero di stelle. In particolare ci sono molte recensioni in cui sono presenti tutte (o quasi) opinioni positive, ma il numero di stelle è inferiore a 3. Questo succede perchè spesso gli utenti hanno opinioni positive su molti aspetti, ma rimangono talmente insoddisfatti di uno da marcare la recensione come negativa. Riuscire a rilevare questo particolare tipo di recensioni è molto utile in quanto possono pesare molto sulla media totale e quindi sulla reputazione di un’attività. Non è corretto che un cliente rimasto soddisfatto su molti aspetti dia importanza solamente a uno o ai pochi negativi e contrassegni la recensione come negativa.

Questo quindi è stato un risultato particolarmente importante perchè un’attività avrebbe la possibilità di segnalare le recensioni incoerenti che arrecano un ingiusto danno di immagine.

Un'altro dato molto interessante emerge dal rilevamento degli outliers rappresentato nella tabella sotto che mostra per ogni stella quante recensioni, con la relativa percentuale, hanno almeno un outlier. Ovviamente nelle recensioni incoerenti sono sempre presenti outliers.

Stelle	Recensioni con outlier	percentuale
1	541	54%
2	783	78%
4	563	56%
5	408	40%

Sebbene sia comprensibile una percentuale maggiore per le recensioni di 2 e 4 stelle, anche quelle da 1 e 5 hanno valori molto rilevanti. Ciò significa che, quando si guardano le recensioni, farsi un giudizio affidandosi solamente al numero di stelle risulta spesso superficiale. Quindi è necessario sempre leggerle per aver un'idea ben precisa perchè circa la metà delle volte troveremo qualcosa di positivo anche se hanno una sola stella.

4 Conclusioni

Lo sviluppo di questo progetto ha fatto emergere quanto l'Opinion Mining sia un campo tanto vasto quanto complesso. Tali difficoltà sono la diretta conseguenza di come noi umani esprimiamo opinioni in modo molto articolato e quindi non facilmente modellabile e schematizzabile per renderlo comprensibile ad una macchina. Nonostante l'estrazione delle opinioni abbia avuto una percentuale di riuscita molto buona, le maggiori difficoltà si sono rivelate nella determinazione della polarità e quindi hanno portato ad una scarsa precisione nella classificazione. Malgrado i risultati, un aspetto positivo è stata la possibilità di individuare le recensioni incoerenti con cui può essere costruito un dataset etichettato per eventuali sviluppi futuri. Inoltre è stato dimostrato che è fondamentale leggere sempre le recensioni e non considerare solamente le stelle.

4.1 Sviluppi futuri

Ci sono molti sviluppi futuri che possono essere affrontati prendendo questo lavoro come base di partenza. Essi sono principalmente legati ai limiti e ai problemi riscontrati:

- Miglioramento del calcolo della polarità che quindi affronti parallelamente il problema della disambiguazione
- Ricerca di una tecnica per identificare l'ironia
- Rilevamento di opinioni implicite e aspetti impliciti
- Sviluppo di un metodo per il rimpiazzamento dei pronomi con il relativo sostantivo a cui si riferiscono
- Categorizzazione degli aspetti

Table 1. Extraction Rules

Gli acronimi POS e DEP si riferiscono alle **Spacy Annotation**

DEP acronyms:

- amod: adjectival modifier
- compound: composition
- acomp: adjectival complement
- advmod: adverbial modifier
- neg: negation
- attr: attribute
- conj: conjunction
- nsubj: nominal subject
- dobj: direct object

POS acronyms:

- NOUN: noun
- PROPN: proper noun
- JJ: adjective
- JJS: adjective superlative

Per ogni relazione $\text{POS}(\text{Adj}) \in \{JJ, JJS\}$ e $\text{POS}(\text{Noun}) \in \{NOUN, PROPN\}$

Relation	DEP	is_Shifted	Modifier	Output
Rel 1 $\text{Adj} \rightarrow \text{Dep} \rightarrow \text{Noun}$	$\text{Dep} \in \{\text{amod}, \text{compound}\}$	True if $\exists \text{Word} \rightarrow \text{Dep} \rightarrow \text{Adj}$ s.t. $\text{Dep} \in \{\text{neg}\}$ \vee $\exists \text{Word} \rightarrow \text{Dep} \rightarrow \text{Noun}$ s.t. $\text{Dep} \in \{\text{neg}\}$ \vee $\exists \text{Noun} \rightarrow \text{Dep} \rightarrow \text{Verb}$ s.t. $\text{Dep} \in \{\text{attr}, \text{dobj}\}$ and $\exists \text{Word} \rightarrow \text{Dep} \rightarrow \text{Verb}$ s.t. $\text{Dep} \in \{\text{neg}\}$	M if $\exists \text{M} \rightarrow \text{Dep} \rightarrow \text{Adj}$ s.t. $\text{Dep} \in \{\text{advmod}\}$	Noun $\text{M} + \text{Adj}$ is_Shifted
Rel 1 Noun Propagation $\text{Noun} \rightarrow \text{Dep2} \rightarrow \text{Noun2}$	$\text{Dep2} \in \{\text{conj}\}$	is_Shifted	M	Noun2 $\text{M} + \text{Adj}$ is_Shifted
Rel 1 Adj Propagation $\text{Adj} \rightarrow \text{Dep2} \rightarrow \text{Adj2}$	$\text{Dep2} \in \{\text{conj}\}$	is_Shifted	M2 if $\exists \text{M2} \rightarrow \text{Dep} \rightarrow \text{Adj2}$ s.t. $\text{Dep} \in \{\text{advmod}\}$	Noun $\text{M2} + \text{Adj2}$ is_Shifted
Rel 2 $\text{Noun} \rightarrow \text{N-Dep} \rightarrow \text{Verb}$ \wedge $\text{Adj} \rightarrow \text{A-Dep} \rightarrow \text{Verb}$	$\text{N-Dep} \in \{\text{nsubj}\}$ $\text{A-Dep} \in \{\text{acom}, \text{attr}\}$	True if $\exists \text{Word} \rightarrow \text{Dep} \rightarrow \text{Adj}$ s.t. $\text{Dep} \in \{\text{neg}\}$ \vee $\exists \text{Word} \rightarrow \text{Dep} \rightarrow \text{Verb}$ s.t. $\text{Dep} \in \{\text{neg}\}$	M if $\exists \text{M} \rightarrow \text{Dep} \rightarrow \text{Adj}$ s.t. $\text{Dep} \in \{\text{advmod}\}$	Noun $\text{M} + \text{Adj}$ is_Shifted
Rel 2 Noun Propagation $\text{Noun} \rightarrow \text{Dep2} \rightarrow \text{Noun2}$	$\text{Dep2} \in \{\text{conj}\}$	is_Shifted	M	Noun2 $\text{M} + \text{Adj}$ is_Shifted
Rel 2 Adj Propagation $\text{Adj} \rightarrow \text{Dep2} \rightarrow \text{Adj2}$	$\text{Dep2} \in \{\text{conj}\}$	is_Shifted	M if $\exists \text{M} \rightarrow \text{Dep} \rightarrow \text{Adj2}$ s.t. $\text{Dep} \in \{\text{advmod}\}$	Noun $\text{M} + \text{Adj2}$ is_Shifted

References

1. Liu, Bing. "Sentiment analysis and opinion mining." *Synthesis lectures on human language technologies* 5.1 (2012): 1-167.
2. Baccianella, Stefano, Andrea Esuli, and Fabrizio Sebastiani. "Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining." *Lrec*. Vol. 10. No. 2010. 2010.
3. Munezero, Myriam, et al. "Are they different? Affect, feeling, emotion, sentiment, and opinion detection in text." *IEEE transactions on affective computing* 5.2 (2014): 101-111.
4. Yelp Open Dataset, <https://www.yelp.com/dataset/documentation/main>