

# Sentiment Analysis: applicazione in un dominio psico-forense

Francesca Guglielmi<sup>1</sup> and Pierpaolo Basile<sup>2</sup> and Antonietta Curci<sup>1</sup> and Giovanni Semeraro<sup>2</sup>

<sup>1</sup>Dipartimento di Scienze della Formazione, Psicologia, Comunicazione, Università di Bari Aldo Moro

<sup>2</sup>Dipartimento di Informatica, Università di Bari Aldo Moro

<sup>1</sup>{francesca.guglielmi, antonietta.curci}@uniba.it

<sup>2</sup>{pierpaolo.basile, giovanni.semeraro}@uniba.it

## Abstract

**English.** This study aims to apply the sentiment analysis to a psycho-forensic context through the analysis of narrative transcriptions related to the crimes committed by violent detainees. We investigate the presence of psychopathy through the Psychopathic Personality Inventory-Revised (PPI-R). Psychopathy is a personality disorder that is characterized by emotional detachment, the lack of empathy and it is often found in the prison population for the brutality of committed crimes. Our study explores possible associations between psychopathy and emotional content present in the narratives. Results show a neutral polarity for both psychopathic and not psychopaths offenders, however it is possible to identify significant emotional characteristics that diversify the narrative transcriptions.

**Italiano.** Il presente studio ha l'obiettivo di applicare la sentiment analysis ad un contesto psico-forense attraverso l'analisi di resoconti narrativi relativi ai crimini commessi da detenuti, autori di reati violenti. Attraverso lo strumento Psychopathic Personality Inventory-Revised (PPI-R), è stata indagata la presenza di psicopatia, disturbo di personalità che si caratterizza per il distacco emotivo, l'assenza di empatia e lo stile di vita antisociale riscontrabile nella popolazione carceraria per l'effettività dei crimini commessi, al fine di esplorare eventuali associazioni tra il disturbo e i contenuti emotivi presenti nelle narrazioni. Dai risultati emerge un tono

*affettivo neutro sia per gli offender psicopatici sia per gli offender non psicopatici, tuttavia è possibile individuare peculiarità emotive significative che diversificano i resoconti narrativi.*

## 1 Introduzione

La psicopatia è un disturbo di personalità caratterizzato da deficit emotivi, stile di vita antisociale, mancanza di senso di colpa ed empatia. Negli psicopatici la coloritura affettiva che generalmente serve a differenziare gli eventi emotivi da quelli neutri non esiste nella stessa misura degli altri individui (Cleckley, 1941).

Gli psicopatici hanno un'attenuata risposta emotiva agli eventi di cui fanno esperienza e nei confronti delle persone con cui interagiscono. I tratti caratterizzanti la psicopatia, si riscontrano in alcuni offender detenuti (Serin, 1996), proprio per le peculiarità relative alla violenza degli atti esperiti da queste persone (Yang et al., 2010). Inoltre, la callosità emotiva, tratto caratterizzante il disturbo, pare possa essere associata ad una percezione minimizzata dell'esperienza emotiva (positiva o negativa) dell'evento vissuto, relativa sia al momento in cui si è verificato sia al richiamo dell'evento stesso (Osumi and Ohira, 2010).

L'incapacità nell'elaborare e nel fare esperienza del contenuto affettivo degli stimoli e degli eventi può avere delle conseguenze sulle capacità di memoria e quindi sulla possibilità di elaborare e di recuperare l'informazione emotiva relativa all'evento esperito. Evidenze empiriche (Burrow et al., 2014) sostengono che le persone psicopatiche soffrono di un deficit emotivo che può avere un impatto sul loro modo di percepire il ricordo degli eventi. Alcuni clinici e ricercatori sostengono

che gli psicopatici generalmente non sono in grado di apprendere il significato emotivo degli eventi; pertanto si ipotizza che i racconti degli offender psicopatici siano scarni di dettagli emotivi rispetto alle narrazioni degli offender non psicopatici (Glass and Newman, 2006; Wilson et al., 2008; Christianson et al., 1996)

La ricerca in questi ambiti necessita di essere implementata, infatti negli ultimi anni la ricerca di base in psicologia cognitiva si è ampliata mediante l'adattamento delle procedure, delle variabili e il coinvolgimento di altri domini scientifici, tra cui la psicolinguistica. Tali adattamenti consentono di allargare le conoscenze sulle caratteristiche di personalità e le loro influenze sui comportamenti in contesti diversificati, come quello penitenziario, per rendere fruibile la messa in atto di programmi di intervento mirati e adatti a supportare persone con tali disturbi. Il presente studio è il risultato di queste integrazioni di domini concettuali e di ricerca, nella fattispecie si applica la sentiment analysis, una metodologia per l'elaborazione automatica del testo al fine di estrarre informazioni soggettive in un contesto psico-forense. A 48 offender detenuti, autori di reati violenti, è stato chiesto, attraverso un compito di free-recall, di narrare il ricordo autobiografico relativo all'evento criminoso commesso per il quale sono stati condannati, tale ricordo è stato trascritto dagli sperimentatori simultaneamente alla narrazione. Inoltre, a ciascun partecipante è stato somministrato il Psychopathic Personality Inventory-Revised (PPI-R) (Lilienfeld and Widows, 2005) per analizzare la presenza di psicopatia. Ciascun resoconto narrativo è stato successivamente analizzato con la metodologia della sentiment analysis attraverso un algoritmo descritto in dettaglio nel paragrafo 2. I risultati di questa analisi sono riportati nel paragrafo 3.

## 2 Metodologia

La metodologia da noi proposta si compone di cinque distinte fasi:

1. reperimento dei resoconti narrativi. Ad un campione composto da 48 uomini ( $M_{età} = 38,69$ ;  $D_s = 8,5$ ) autori di reati violenti, detenuti presso l'Istituto penitenziario di Turi (BA) e la Casa circondariale di Trani (BA) è stato chiesto di narrare, attraverso un compito di free-recall, il ricordo autobiografico relativo al crimine per il quale sono stati condan-

nati. Le narrazioni sono state trascritte dagli sperimentatori in simultanea;

2. rilevazione della presenza/assenza di psicopatia misurata attraverso il PPI-R (Lilienfeld and Widows, 2005) che valuta i tratti di personalità del soggetto attraverso l'indagine di quei comportamenti e peculiarità cognitive, emotive e percettive che costituiscono i tratti tipici della psicopatia. Il test è composto da 154 item, da cui si ricavano 8 scale di contenuto, aggregate in 3 fattori sovraordinati: *Dominanza priva di paura*, *Impulsività auto-centrata* e *Freddezza emotiva*. Ogni item prevede 4 alternative di risposta (vero, abbastanza vero, abbastanza falso, falso). Le somministrazioni si sono tenute previa autorizzazione del P.R.A.P. della Regione Puglia e previa autorizzazione individuale all'uso dei dati personali nei limiti stabiliti dalla legge. Le somministrazioni si sono svolte in una cella protetta, alla presenza di due somministratori e una guardia penitenziaria;
3. elaborazione dei resoconti narrativi, come descritto nel paragrafo 2.1, per l'estrazione automatica di attributi emotivi e di sentimento;
4. analisi dei risultati per l'individuazione di una correlazione tra i punteggi del PPI-R e gli attributi estratti dall'analisi automatica del testo;
5. analisi degli attributi estratti dall'elaborazione del testo con tecniche di apprendimento automatico.

La metodologia proposta ha l'obiettivo di verificare se l'analisi automatica del testo può rilevare alcune caratteristiche emozionali o di sentimento associabili alla psicopatia nei detenuti che hanno commesso crimini particolarmente violenti.

### 2.1 Elaborazione del testo

Ogni trascrizione è stata processata da una pipeline di strumenti per l'analisi automatica del linguaggio per la lingua italiana. In particolare, ogni trascrizione è sottoposta alle seguenti operazioni:

1. **Tokenizzazione e PoS-tagging:** il testo è suddiviso in token (parole) e ad ogni token è associato un tag che identifica la parte del discorso (sostantivo, aggettivo, verbo, avverbio, preposizione, articolo, ...). Queste operazioni sono realizzate attraverso il tool

open-source Apache OpenNLP<sup>1</sup>;

2. **Lemmatizzazione:** sfruttando le informazioni sul PoS-tag, riconduce ogni parola al suo lemma. Questa operazione è necessaria per poter accedere a dizionari elettronici, ad esempio quelli utilizzati nelle operazioni successive per l'analisi del sentiment. L'operazione è stata realizzata attraverso l'utilizzo della risorsa linguistica Morph-it!<sup>2</sup> (Zanchetta and Baroni, 2005);
3. **Analisi del sentiment:** l'analisi del sentiment prevede l'associazione di uno score di polarità ad ogni parola presente nella trascrizione, ed il calcolo di uno score di sentiment sull'intera trascrizione. Lo score di sentiment sull'intera trascrizione è calcolato seguendo due differenti approcci: 1) come media degli score di sentiment associati ad ogni parola; 2) come differenza tra gli score massimi di polarità positiva e negativa. Questo processo è realizzato utilizzando la risorsa linguistica Sentix (Basile and Nissim, 2013). Inoltre, la polarità dei termini è invertita se compaiono in un contesto di negazione. Il contesto di negazione è individuato dalla parola "non" e si estende fino all'individuazione di uno dei seguenti caratteri {., :;!};
4. **Analisi delle emozioni:** ad ogni parola sono associate delle etichette emotive utilizzando la risorsa linguistica WordNet-Affect (Strapparava and Valitutti, 2004).

L'analisi del sentiment e quella delle emozioni necessitano di maggiori dettagli implementativi. Poiché le risorse linguistiche utilizzate sono dei dizionari basati sul significato si rende necessaria una strategia per poter accedere a questi dizionari attraverso il lemma. Un lemma può avere più significati; nel caso di lemma polisemico si è scelto il significato più frequente. La frequenza dei significati per la lingua italiana è calcolata utilizzando la risorsa MultiSemCor (Bentivogli and Pianta, 2005). Lo stesso procedimento viene applicato a WordNet-Affect in presenza di lemmi polisemici.

## 2.2 Analisi con approcci di apprendimento automatico

Al fine di analizzare l'associazione tra gli attributi linguistici e la psicopatologia, abbiamo formalizzato

<sup>1</sup><https://opennlp.apache.org/>

<sup>2</sup><http://sslmitdev-online.sslmit.unibo.it/linguistics/morph-it.php>

un problema di apprendimento di tipo binario in cui ogni racconto è classificato come *psicopatico* o *non psicopatico* in base ai risultati del PPI-R. Ogni racconto è descritto da una serie di attributi: polarità media, score massimo di polarità positiva, score massimo di polarità negativa, differenza tra gli score massimi di polarità e, per ogni classe emotiva prevista da WordNet-Affect, numero di parole che ricadono in tale classe.

Lo scopo di questa analisi non è quello di misurare la bontà predittiva degli attributi estratti dal testo, ma di capire il potere discriminante di ogni attributo. Per questo motivo si è scelto l'algoritmo C4.5 (Quinlan, 1993) basato sugli alberi di decisione. In particolare l'algoritmo genera un albero di decisione basandosi sul concetto di entropia dell'informazione.

Ad ogni nodo dell'albero, l'algoritmo C4.5 sceglie l'attributo che divide più efficacemente i dati in una o nell'altra classe. Il criterio di suddivisione è il guadagno d'informazione (differenza di entropia). L'attributo con il più alto guadagno d'informazione viene scelto per prendere la decisione. L'algoritmo C4.5 viene applicato in maniera ricorsiva ai nuovi sottoinsiemi di dati che si vengano a creare ad ogni passo.

L'idea è che l'albero di decisione possa darci un'indicazione di quali siano gli attributi più discriminanti per l'individuazione dei racconti appartenenti a soggetti affetti da psicopatologia.

## 3 Analisi sperimentale

Il punteggio totale di psicopatologia, che ha consentito la discriminazione dei due gruppi: offender psicopatici (n. 20) e offender non psicopatici (n. 28), è stato calcolato attraverso moduli di autoscoring sommando, per ciascun soggetto, i punteggi ottenuti nei tre fattori (*Dominanza priva di paura*, *Impulsività auto centrata* e *Freddezza emotiva*). Tale punteggio grezzo è stato poi convertito in punti *t*, quindi standardizzato. Attraverso il t-test, test statistico parametrico che permette di verificare se il valore medio di una distribuzione si discosta significativamente da un certo valore di riferimento ad un livello di significatività inferiore a 0,05, emerge che sia i punteggi totali di Psicopatologia ( $t(46) = -9,86$ ), sia dei tre fattori: *Dominanza priva di paura* ( $t(46) = -8,24$ ); *Impulsività autocentrata* ( $t(46) = -9,86$ ) e *Freddezza emotiva* ( $t(46) = -8,24$ ) hanno delle differenze significative con valori di  $p < 0,05$ . Questo risultato la-

Tabella 1: Correlazione di Pearson tra gli indicatori della Sentiment Analysis e i fattori del PPI-R nel gruppo di offender psicopatici (n = 20). (\*) p < 0.05

Sentiment Analysis	Psicopatia: PPI-R			
	Impulsività autocentrata	Dominanza priva di paura	Freddezza emotiva	Psicopatia totale
Polarità				
$M_{pol}$	-0,260	0,103	0,097	-0,095
$Max_{pos}$	-0,006	0,507*	0,335	0,305
$Max_{neg}$	-0,116	0,104	-0,121	-0,054
$Max_{pos} - Max_{neg}$	0,069	0,291	0,311	0,249

scia ragionevolmente supporre che i due gruppi si differenziano per i tratti caratterizzanti la psicopatia relativi allo stile di vita antisociale, la callosità emotiva e le caratteristiche legate alla mancanza di paura, gestione dello stress e influenza sociale.

Successivamente, i resoconti narrativi sono stati analizzati mediante tecniche di elaborazione del linguaggio naturale al fine di estrarre attributi emotivi e di polarità del sentimento (*positiva*, *negativa* e *neutra*). Gli attributi così estratti sono stati utilizzati per impostare il problema di apprendimento automatico descritto nel paragrafo 2.2. Per realizzare ciò è stata utilizzata l'implementazione J48 dell'algoritmo C4.5 presente nel tool open-source Weka<sup>3</sup> (Hall et al., 2009).

### 3.1 Risultati

Per analizzare l'associazione tra psicopatia e punteggi ottenuti negli score della Sentiment Analysis abbiamo utilizzato analisi correlazionali attraverso le quali è emersa una correlazione significativa tra  $Max_{pos}$  e il fattore *Dominanza priva di paura* del PPI (Tabella 1) nel gruppo composto da offender psicopatici. Questo risultato sembra sostenere l'ipotesi che la freddezza emotiva dello psicopatico predisponga ad una ricerca attiva di tutte quelle situazioni e sensazioni forti ed eccitanti che possano in qualche modo indurre uno stato di eccitazione in loro (Hervé et al., 2007), come il commettere i crimini violenti di cui si sono resi protagonisti e la loro tendenza a correre rischi (Ellis, 1987). Al contrario nel gruppo composto da offender non psicopatici non è emersa alcuna correlazione significativa.

Per quanto riguarda l'output fornitoci dall'albero di decisione, si evince che la differenza di polarità è l'attributo più discriminante, in grado di riconoscere da solo quasi il 50% degli individui

affetti da psicopatia. Ancora più interessante è il valore per il quale avviene la suddivisione del campione. In particolare, se la differenza di polarità è maggiore di -0,125, l'individuo viene classificato come affetto da psicopatia. Questo risultato è molto rilevante in quanto la differenza di polarità nel campione assume valori nell'intervallo [-0,625; 0,25]. Questo significa che le narrazioni neutre, che hanno una differenza vicina allo zero, risultano scritte da soggetti affetti da psicopatia, in linea con l'ipotesi che gli psicopatici abbiano un ricordo povero di dettagli emotivi che lascia ragionevolmente presupporre un'attenuata risposta emozionale agli eventi.

È anche importante sottolineare come la polarità sia sbilanciata verso la negatività. Ciò suggerisce che, in assenza di psicopatia, il tipo di ricordo associato all'evento criminoso risulta essere negativo. Continuando ad analizzare l'albero di decisione si scopre che i nodi successivi si basano sugli attributi di polarità massima positiva e negativa, e solo uno dei nodi fa riferimento all'etichetta "piacere" di WordNetAffect.

## 4 Conclusioni

In questo lavoro abbiamo cercato di capire se i risultati prodotti da un algoritmo di sentiment analysis possano in qualche modo essere associati alla psicopatia. In particolare, abbiamo analizzato la presenza di psicopatia, in un campione di offender detenuti poichè i tratti caratterizzanti di tale disturbo, distacco emotivo, mancanza di senso di colpa, stile di vita antisociale, è possibile riscontrarli negli offender detenuti per le peculiarità relative alla violenza degli atti esperiti e allo stile di vita condotto.

I risultati ottenuti dalla sentiment analysis sono stati confrontati con quelli ottenuti attraverso il test PPI-R che valuta i tratti di personalità del sog-

<sup>3</sup><http://www.cs.waikato.ac.nz/ml/weka/>

getto attraverso l'indagine di quei comportamenti e peculiarità cognitive, emotive e percettive che costituiscono i tratti tipici della psicopatia.

I risultati ottenuti attraverso l'utilizzo di un approccio basato sull'induzione di alberi di decisione dimostrano una dipendenza tra la presenza di psicopatia e la differenza tra la massima polarità positiva e quella negativa individuate nella narrazione del detenuto. Inoltre, è stata riscontrata una correlazione tra lo score massimo di polarità e il fattore *Dominanza priva di paura* attraverso l'utilizzo di test statistici.

I risultati ottenuti sono incoraggianti e dimostrano l'efficacia di ampliare la ricerca di base in psicologia cognitiva con strumenti innovativi come quello della sentiment analysis e in genere dell'analisi automatica del testo. Lo studio qui riportato, seppur rilevante, è comunque da considerarsi preliminare per due ragioni: la ridotta dimensione del campione e l'utilizzo di tecniche abbastanza semplici per l'analisi del sentiment.

## Acknowledgement

Questo lavoro è parzialmente supportato dal progetto "Multilingual Entity Liking" finanziato dalla Regione Puglia con il programma FutureInResearch.

## References

- Valerio Basile and Malvina Nissim. 2013. Sentiment analysis on Italian tweets. In *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 100–107, Atlanta.
- Luisa Bentivogli and Emanuele Pianta. 2005. Exploiting parallel texts in the creation of multilingual semantically annotated resources: the Multi-SemCor Corpus. *Natural Language Engineering*, 11(03):247–261.
- Ashley N Burrow, Nichole Currence, Diana Lemus, Amber E DeBono, Matthew T Crawford, and W Richard Walker. 2014. Psychopaths view autobiographical memories as less memorable, important, and emotional than normal individuals. *International Journal of Humanities and Social Science*, 4:1–9.
- Sven-Åke Christianson, Adelle E Forth, Robert D Hare, Catherine Strachan, Lars Lidberg, and Lars Håkan Thorell. 1996. Remembering details of emotional events: A comparison between psychopathic and nonpsychopathic offenders. *Personality and Individual Differences*, 20(4):437–443.
- H Cleckley. 1941. *The mask of sanity*. st. louis. *MO: Mosby*.
- Lee Ellis. 1987. Relationships of criminality and psychopathy with eight other apparent behavioral manifestations of sub-optimal arousal. *Personality and Individual Differences*, 8(6):905–925.
- Samantha J Glass and Joseph P Newman. 2006. Recognition of facial affect in psychopathic offenders. *Journal of abnormal psychology*, 115(4):815.
- Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H Witten. 2009. The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18.
- Hugues Hervé, Barry S Cooper, and JOHN C Yuille. 2007. Memory formation in offenders: Perspectives from a biopsychosocial model of eyewitness memory. *Offenders' memories of violent crimes*, pages 38–74.
- Scott O Lilienfeld and Michelle R Widows. 2005. *PPI-R: Psychopathic personality inventory revised: Professional Manual*. Psychological Assessment Resources, Incorporated.
- Takahiro Osumi and Hideki Ohira. 2010. The positive side of psychopathy: Emotional detachment in psychopathy and rational decision-making in the ultimatum game. *Personality and individual differences*, 49(5):451–456.
- J Ross Quinlan. 1993. *C4. 5: programs for machine learning*. Morgan Kaufmann Publishers.
- Ralph C Serin. 1996. Violent recidivism in criminal psychopaths. *Law and Human Behavior*, 20(2):207.
- Carlo Strapparava and Alessandro Valitutti. 2004. WordNet Affect: an Affective Extension of WordNet. In *LREC*, volume 4, pages 1083–1086.
- Kevin Wilson, Sabrina Demetriooff, and Stephen Porter. 2008. A pawn by any other name? social information processing as a function of psychopathic traits. *Journal of research in personality*, 42(6):1651–1656.
- Yaling Yang, Adrian Raine, Patrick Colletti, Arthur W Toga, and Katherine L Narr. 2010. Morphological alterations in the prefrontal cortex and the amygdala in unsuccessful psychopaths. *Journal of abnormal psychology*, 119(3):546.
- Eros Zanchetta and Marco Baroni. 2005. Morph-it! A free corpus-based morphological resource for the Italian language. *Corpus Linguistics 2005*, 1(1).