

# Le nuove frontiere metodologiche nell'era dei big data: l'Emotional Text Mining

[eyesreg.it/2020/le-nuove-frontiere-metodologiche-nellera-dei-big-data-lemotional-text-mining/](https://eyesreg.it/2020/le-nuove-frontiere-metodologiche-nellera-dei-big-data-lemotional-text-mining/)

**di:** *Francesca Greco*

EyesReg, Vol.10, N.4, Luglio 2020. Numero Speciale: “Nuovi orizzonti di ricerca per le Scienze Regionali”

Negli ultimi venti anni si è assistito alla crescente diffusione di internet, che ha permeato molte delle aree della nostra vita quotidiana. Nel recente periodo di confinamento dovuto alla pandemia da covid-19, internet ha consentito a milioni di persone di navigare sul web, cercare e condividere informazioni quotidianamente. Alcune delle attività lavorative sono continuate grazie al telelavoro e si è potuta conservare, almeno in parte, una vita sociale.

Il costante aumento dello scambio di informazioni attraverso le piattaforme dei social media ha anche reso disponibile una grande quantità di dati, che rappresentano una fonte inestimabile di informazioni per tutti coloro che sono interessati a conoscere le preferenze, le opinioni e i sentimenti delle persone. Per questo motivo, i social media ed i siti di social network, come Facebook e Twitter, in quanto snodi principali di raccolta/osservazione dei comportamenti digitali dei propri utenti, hanno iniziato ad assumere un ruolo crescente nella politica del mondo reale.

Nell'era dei big data (1), quindi, il problema non è come raccogliere determinate informazioni, ma come *estrarle*, vale a dire quale metodo utilizzare per trasformare i dati in conoscenza attraverso l'analisi di dati già rilevati, costantemente prodotti dall'attività dell'uomo, eterogenei e non strutturati nel 95% dei casi. La costante raccolta di queste informazioni, infatti, genera una tale mole di dati da determinare un problema di trasformazione delle informazioni in conoscenza, ponendo una sfida sia in termini di avanzamento tecnologico che metodologico. Per avere un'idea del problema basti pensare che, all'inizio del 2017, Twitter contava oltre 300 milioni di utenti che producevano una media di circa 500 milioni di messaggi in diverse lingue ogni giorno.

## **L'analisi dei big data**

Negli ultimi anni sono state sviluppate un numero crescente di tecniche di analisi, che sono diventate degli strumenti utili, veloci ed economici, rispetto alla rilevazione tradizionale, per l'esplorazione dei fenomeni sociali. Ad esempio, la comunicazione online è analizzata attraverso tecniche di *text mining* per diversi scopi come: la pianificazione dei prodotti da mettere sul mercato, la previsione del comportamento di voto, la gestione delle catastrofi naturali, il monitoraggio delle campagne promozionali, la valutazione di siti web, la valutazione dei feedback dei clienti nel marketing digitale.

Tra i metodi proposti, l'analisi del *sentiment* sta avendo un crescente successo poiché consente di identificare e misurare un'opinione, una convinzione o un giudizio personale, che non si basa sul ragionamento razionale, ma sull'emozione soggettiva. Questo aspetto è rilevante poiché molte delle scelte che le persone fanno sono dettate o, comunque, influenzate da questo aspetto. Ad esempio, le aziende possono essere interessate a comprendere cosa indurrà i potenziali consumatori ad acquistare un determinato prodotto, così come un leader politico può essere interessato a comprendere quali temi stanno a cuore agli elettori.

L'utilizzo del *text mining* per misurare il *sentiment* è molto diffuso e sostanzialmente si basa su due approcci:

- quello semantico, in cui le parole vengono considerate solo per il loro significato esplicito
- quello semiotico, in cui le parole vengono considerate come simboli dei modi di pensare delle persone.

Nel primo caso, spesso, si ricorre a procedure che misurano il *sentiment* di un testo attraverso l'utilizzo di dizionari o di training set precodificati. Nel secondo caso, invece, la misurazione viene fatta successivamente alla classificazione e si basa sul principio dell'associazione delle parole.

Questo approccio offre l'innegabile vantaggio di tenere in considerazione le caratteristiche contestuali della comunicazione, vale a dire che una parola assume un significato, positivo o negativo, in relazione alle parole a cui è associata. Ad esempio, la parola *bomba* di solito implica un sentimento negativo, come nella frase "La bomba è esplosa nella metropolitana", tuttavia essa può assumere un sentimento positivo di ammirazione, come nella frase "quella ragazza è una bomba!". Pertanto, la presenza o l'assenza di parole con un *sentiment* specifico in una frase non implica necessariamente la possibilità di classificarlo e questa è la criticità dell'approccio semantico.

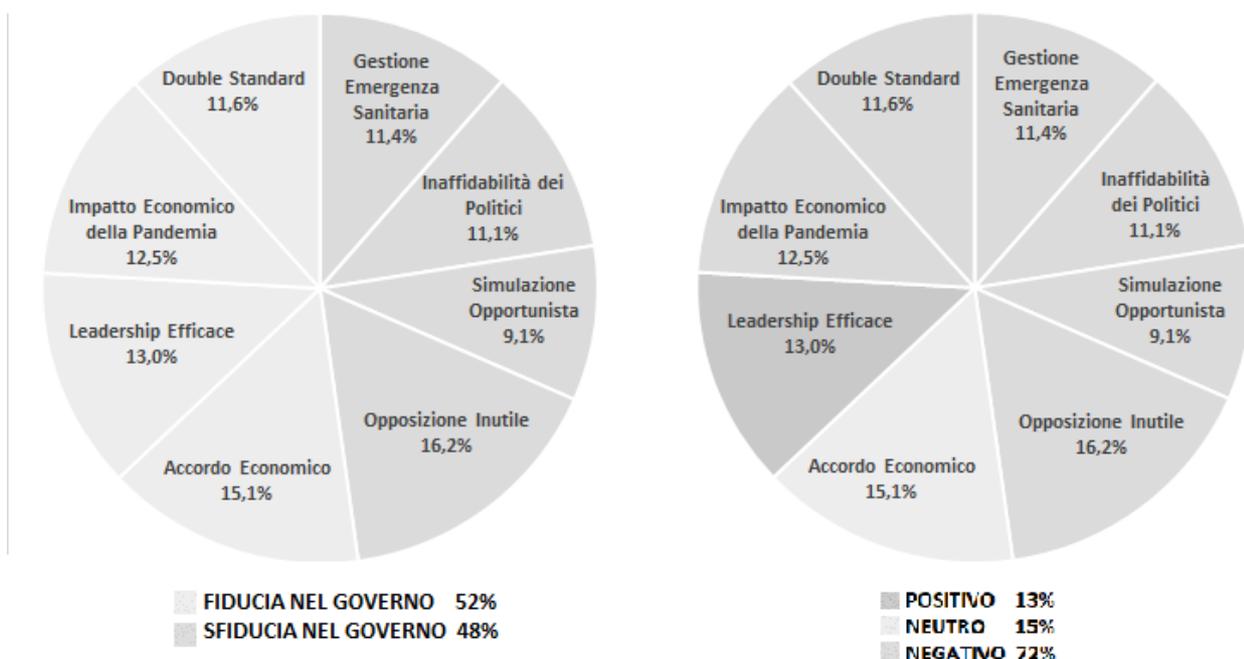
L'approccio semantico, peraltro, offre il vantaggio non indifferente di connettere la comunicazione con il comportamento, effettuando di fatto una profilazione a partire dalla comunicazione. Vi è, infatti, una stretta connessione tra le parole e i comportamenti anche se questa relazione non è diretta. Come sosteneva il pubblicitario David Ogilvy, le persone non sempre fanno ciò che sentono, né dicono ciò che pensano e, soprattutto, non si comportano coerentemente con ciò che dicono. Tuttavia, esistono delle procedure, come l'ETM (Emotional Text Mining) che consentono di superare questo problema essendo una tecnica di profiling sociale. Sostanzialmente, ciò che si guarda non è il contenuto della comunicazione ma la sua modalità, vale a dire quali parole vengono usate e come vengono associate poiché, come è noto in psicologia, ciò è espressione del funzionamento mentale che determina anche il comportamento.

L'ETM, quindi, analizza il testo per inferire il funzionamento mentale che non determina solo la comunicazione, ma anche gli atteggiamenti, le aspettative e i comportamenti delle persone, ed è per questo motivo che si può parlare di tecnica di

profilazione. Mentre il funzionamento mentale procede dal livello semiotico a quello semantico nel generare il testo, la procedura statistica simula il processo inverso del funzionamento mentale, dal livello semantico a quello semiotico. Per questo motivo, ETM esegue una sequenza di procedure di sintesi, dal pretrattamento del testo e la selezione delle parole chiave all'analisi multivariata, per identificare il livello semiotico, partendo da quello semantico. È una procedura automatica, veloce e relativamente semplice, che può essere utilizzata per estrarre informazioni in merito a quelle che potremmo definire le *opinioni emotivamente guidate*. A differenza di altre tecniche di sentiment analysis, l'EMT non solo misura il *sentiment*, ma fornisce anche informazioni qualitative sullo stesso.

Ad esempio, in uno studio volto a identificare la percezione pubblica dei leader di governo e dell'opposizione, nel periodo del confinamento (fase 1) (Greco & La Rocca, 2020), emerge, come si può facilmente intuire, un *sentiment* prevalentemente negativo (72%). Tuttavia, l'ETM ha consentito di individuare otto differenti temi prevalentemente connessi con l'emergenza sanitaria: *Gestione dell'Emergenza Sanitaria*, *Inaffidabilità dei Politici*, *Simulazione Opportunista*, *Opposizione Inutile*, *Accordo Economico Internazionale*, *Leadership Efficace*, *Impatto Economico della Pandemia* e *Double Standard*, vale a dire l'utilizzo di due pesi e due misure nel valutare il livello di democrazia in Italia e in altri paesi. Ciò consente di rimisurare il *sentiment* in relazione al livello di fiducia/sfiducia nel Governo, ottenendo dei risultati radicalmente diversi, come si può osservare confrontando i due grafici.

Figura 1. Temi emergenti del *sentiment* in Italia durante l'emergenza sanitaria. Fonte: elaborazione dell'autrice.



Come si può facilmente intuire, questa informazione può essere utile, ad esempio, a orientare i processi comunicativi in modo efficace. Infatti, l'ETM identifica un set di parole connesse al tema e allo specifico *sentiment* che possono essere utilizzate per

organizzare la comunicazione. Inoltre, lo studio evidenzia come le strategie di comunicazione adottate dai leader politici, che negli anni passati sono risultate efficaci, durante la pandemia sono state controproducenti spesso sortendo l'effetto opposto a quello desiderato e diminuendo la visibilità del leader in termini di messaggi ad esso concernenti. Come sosteneva Oscar Wilde: "C'è una sola cosa al mondo peggiore del far parlare di sé, ed è il non far parlare di sé."

Per le caratteristiche fin qui descritte, l'ETM è una procedura piuttosto versatile, non vincolata alla disponibilità di un dizionario precodificato del *sentiment*. Per questo motivo, è stata utilizzata in diversi contesti e per diversi scopi: l'analisi delle campagne elettorali per la valutazione degli esiti in termini di voto; la misurazione la percezione pubblica della sicurezza per implementare la misura della qualità della vita; lo studio della percezione pubblica della crisi di governo e di quella industriale del mezzogiorno dello scorso anno; il confronto tra paesi diversi per quanto riguarda i motivi che favoriscono la donazione degli organi, le pratiche di procreazione assistita o la formazione e l'inserimento lavorativo delle persone diversamente abili; il brand management, con l'identificazione delle comunità di acquisto in base alla rappresentazioni del brand e l'identificazione di specifici comportamenti di interazione online; la valutazione della formazione e altro ancora.

Di qui anche l'utilità per le scienze regionali, soprattutto nel momento in cui l'analisi del *sentiment* viene verificata rispetto alla variabile dell'appartenenza a una specifica area geografica o regione. Ad esempio, in uno studio recente si è indagato l'impatto che le vicende di cronaca relative allo stabilimento siderurgico ex-Ilva di Taranto hanno avuto sulla percezione pubblica della sicurezza anche a livello nazionale e non solo locale, in modo particolare, la sicurezza economica, lavorativa, sanitaria e ambientale (Greco & Polli, 2020). Lo studio dimostra come si possa parlare di *glocal sentiment*, poiché eventi che accadono in una specifica area geografica, inevitabilmente, hanno un impatto anche a livello nazionale o sovranazionale. Conoscere il *sentiment* di specifici eventi locali, quindi, diventa un elemento essenziale per comprendere il complesso intreccio tra il locale e il globale e le sue ricadute pratiche in una specifica area geografica..

*Francesca Greco, La Sapienza Università di Roma*

## **Bibliografia**

Cordella, B., Greco, F., Carlini, K., Greco, A., Tambelli, R. (2018). Infertilità e procreazione assistita: evoluzione legislativa e culturale in Italia. *Rassegna di Psicologia*, 35(3), 45-56. DOI: 10.4458/1415-04 ISSN: 1125-5196

Cordella, B., Greco, F., Meoli, P., Palermo, V. Grasso, M. (2020). Educational culture and job market: A text mining approach. in D.F. Iezzi, D. Mayaffre, M. Misuraca (eds.), *Text Analytics. Advances and Challenges*, Springer, Heidelberg. ISBN 978-3-030-52679-5

Greco, F. (2016). *Integrare la disabilità. Una metodologia interdisciplinare per leggere il cambiamento culturale*. Milano: Franco Angeli, pp.144. ISBN: 9788891740755

Greco, F. (2016). *Les Lois et le changement culturel: Le handicap en Italie et en France*. Roma : Sapienza Università Editrice. ISBN 978-88-9377-044-6

Greco, F. (2019). Il dibattito sulla migrazione in campagna elettorale: Confronto tra il caso francese e italiano. *Culture e Studi nel Sociale*, 4(2), 205-13.  
<http://www.cussoc.it/index.php/journal/article/view/97>

Greco, F. (2020). Rendere visibile l'invisibile: disabilità e sviluppo sostenibile tra indicatori e stereotipi. *Culture e Studi del Sociale*, 5(1). Retrieved  
<http://www.cussoc.it/index.php/journal/issue/archive>

Greco, F., Alaimo, L. Celardo L. (2018). Brexit and Twitter: The voice of people. In D.F. Iezzi, L. Celardo & M. Misuraca (Eds.), *JADT' 18: Proceedings of the 14th International Conference on Statistical Analysis of Textual Data*, pp. 327-334. Roma: Universitalia. ISBN 978-88-3293-137-2

Greco, F., Celardo, L., Alaimo, L.M. (2018). Brexit in Italy: Text Mining of Social Media. In Abbruzzo, A., Piacentino, D., Chiodi, M., & Brentari, E. (Eds.), *Book of short Papers SIS 2018*, pp. 767-772. Milano: Pearson. ISBN-9788891910233

Greco, F., La Rocca, G. (2020). The Topics-scape of the Pandemic Crisis: The Italian sentiment on political leaders. *Culture e Studi del Sociale*, 5(1), Special issue, 131-145.  
<http://www.cussoc.it/index.php/journal/issue/archive>

Greco, F., Maschietti, D. Polli, A. (2017). Emotional text mining of social networks: The French pre-electoral sentiment on migration. *Rivista Italiana di Economia Demografia e Statistica*, 71(2), 125-136. <https://ideas.repec.org/a/ite/iteeco/170205.html>

Greco, F., Monaco, S., Di Trani, M., Cordella, B. (2019). Emotional text mining and health psychology: the culture of organ donation in Spain. In M. Carpita and L. Fabbris (Eds.), *ASA Conference 2019 – Book of Short Papers Statistics for Health and Well-being, University of Brescia, September 25-27, 2019* (pp. 125-129). Padova: CLEUP.

Greco, F., Polli, A. (2019). Anatomy of a government crisis. Political institutions, security, and consensus. In L.S. Alaimo, A. Arcagni, E. di Bella, F. Maggino & M. Trapani (Eds.), *Libro dei Contributi Brevi: AIQUAV 2019, VI Convegno Nazionale dell'Associazione Italiana, per gli Studi sulla Qualità della Vita, Benessere Collettivo e Scelte Individuali, Fiesole (FI), 12-14 Dicembre 2019* (pp. 177-183). Genova: Genova University Press. ISBN: 978-88-94943-76-4

Greco, F., Polli, A. (2019). Vaccines In Italy: The Emotional Text Mining of Social Media. *Rivista Italiana di Economia Demografia e Statistica*, 73(1), 89-98.  
<https://ideas.repec.org/a/ite/iteeco/190106.html>

Greco, F., Polli, A. (2020). Emotional Text Mining: Customer profiling in brand management. *International Journal of Information Management*. 51, 101934. <https://doi.org/10.1016/j.ijinfomgt.2019.04.007>.

Greco, F., Polli, A. (2020). La sicurezza tra percezione pubblica e statistiche ufficiali. In U. Conti and C. Federici (Eds.), *Vivere i territori mediani: identità territoriali, emergenze sociali e rigenerazione dei tessuti urbani* (pp. 37-60). Roma: Meltemi.

Greco, F., Polli, A. (2020). Security Perception and People Well-Being. *Social Indicator Research*, 1-18. <https://doi.org/10.1007/s11205-020-02341-8>

Greco, F., Polli, A. (2020). The political debate on immigration in the election campaigns in Europe. In A. Przegalinska, F. Grippa and P. Gloor (Eds.), *Digital Transformation of Collaboration: Proceedings of the 9th International COINs Conference*. Springer.

Greco, F., Polli, A. (2020a). Crisi industriale nel Mezzogiorno. Glocal sentiment, cittadini e istituzioni. In R. Veraldi (Ed.), *Intersezioni sociologiche sullo sviluppo (locale). Società fuori squadr: Italia a due velocità in un Europa cieca*. Roma: Edizioni Universitarie Romane. ISBN: 978-88-6022-379-1

Laricchiuta, D., Greco, F., Piras, F., Cordella, B., Cutuli, D., Picerni, E., Assogna, F., Lai, C., Spalletta, G. Petrosini, L. (2018). “The grief that doesn’t speak”: Text Mining and Brain Structure. In D.F. Iezzi, L. Celardo & M. Misuraca (Eds.), *JADT’ 18: Proceedings of the 14th International Conference on Statistical Analysis of Textual Data*, pp. 419-427. Roma: Universitalia. ISBN 978-88-3293-137-2

---

## Note

(1) Quando si parla di big data si intende una raccolta di informazioni talmente estesa in termini di volume (quantità), velocità (quantità prodotta in una specifica unità di tempo) ed eterogeneità (es., frequenze, testi, tempi, ecc.) da costituire un problema di estrazione dell’informazione. Il concetto di big data non definisce una quantità specifica di dati e la sua estensione è funzione degli strumenti tecnologici e dei metodi di analisi di cui si dispone.