Fake News e Persuasione Occulta: come riconoscerle, come evitarle (2° Parte)

Data 07 febbraio 2021
Categoria Medicinadigitale

Come EWE classifica le Emozioni
EWE, acronimo di Emotion Word Embedding, è un insieme di algoritmi in grado di identificare il contenuto emozionale di molteplici espressioni linguistiche, che vengono classificate secondo le sei emozioni fondamentali proposte da Paul Ekman(1): rabbia, disgusto, paura, gioia, tristezza, sorpresa.

> Secondo lo psicologo Paul Ekman le emozioni umane possono essere ricondotte a sei fondamentali stati d'animo [b]Gioia, Tristezza, Paura, Rabbia, Sorpresa e Disgusto[/b]. I molteplici, complessi stati d'animo che ogni essere umano esperimenta nella propria vita(2) potrebbero essere ricondotti a queste sei emozioni fondamentali: as esempio la "noia" sarebbe un insieme di tristezza e disgusto,la "invidia" di rabbia e tristezza ecc.
> Molti ricercatori hanno criticato questa classificazione, ma per le finalità che si propongono I "grandi fratelli del web" la classificazione di Elkman, purtroppo, funziona…

EWE fu sviluppato da Ameeta Agrawal(3,4) partendo da numerosissimi documenti prelevati dal web ed analizzati da reti neurali, supervisionate da esperti informatici ,che classificavano le varie espressioni linguistiche sia dal punto di vista qualitativo, inserendole nella griglia delle sei emozioni fondamentali, sia da quello quantitativo, associando ad ogni espressione linguistica un peso ovvero un grado di emozione. Inutile sottolineare come queste operazioni, apparentemente semplici, richiedano in realtà grande familiarità con strumenti matematici raffinati e complessi quali il calcolovettoriale.

Word2vec e la Interpretazione del contesto
I dispositivi di intelligenza artificiale hanno molti pregi ma un grave difetto: non possiedono Il buon senso di cui è dotato ogni essere umano. Ad esempio una frase un po' elaborata quale:"Iniziò a parlare...quale ottima occasione per stare zitto!!!" è chiaramente una proposizione ironica che con una certa finezza esprime una valutazione decisamente negativa sull'oratore.
Una semplice analisi verbale effettuata dagli algoritmi utilizzati nel decennio scorso non troverà alcuna espressione chiaramente negativa e potrà addirittura individuare nel sintagma "ottima occasione" un indicatore di apprezzamento. Tale handicap dei sistemi di intelligenza artificiale è stato in larga parte superato con la creazione Word2vec, che nacque grazie alle intuizioni di Tomas Mikolov(5) di Google.

Word2vec si avvale in particolare di due importanti algoritmi: CBOW, che prevede la parola a partire dal contesto ovvero dalle parole precedenti, e Skip-gram che partendo da parole chiave esamina i diversi contesti nei quali la parola chiave si può situare arrivando successivamente a prevedere le espressioni linguistiche successive. Wod2vec ha consentito un importante salto di qualità nell'analisi automatica da parte delle reti neurali di milioni di testi del web. Anche con questi algoritmi, ovviamente, si debbono utilizzare strumenti matematici di grande complessità quali il calcolo combinatorio, inferenze statistiche, valutazioni probabilistiche, ecc.

EWE e gli algoritmi analoghi classificano le emozioni, Word2vec e gli algoritmi analoghi effettuano analisi del contenuto linguistico dei testi. Qualcuno pensò di integrare i due approcci e fece " bingo"!!! In effetti se i due piani di classificazione vengono integrati e sottoposti ad ulteriori supervisioni critiche la potenza del sistema analitico aumenta considerevolmente.
Ma non è sufficiente: nella foresta del web e dei social vengono analizzate sia le notizie che i commenti alle notizie: notizie e commenti vengono sottoposti ad una analisi multivariata in cui ogni fonte di informazione utile viene estratta,registrata, catalogata, e messa in relazione con le altre fonti effettuando una valutazione quantitativa, qualitativa ecomparativa…

I "Like", questi insidiosi seduttori
Il like, ovvero "mi piace" o, per esteso, "sono d'accordo", è un commento spontaneo quanto mai diffuso nei media. Cosa si potrebbe dedurre da un semplice like? Apparentemente nulla o comunque qualcosa di poco conto. ..
Nulla di più errato: dall'insieme dei "like" espressi è possibile estrarre profili psicologici affidabili e notizie sufficientemente precise sugli orientamenti ideologici, politici e religiosi di chi scrive.
Di converso chi si muove nella legalità ed ha a cuore i diritti delle persone, come il gruppo di Luca de Alfaro, dall'analisi della diffusione dei like può individuare con una precisione attorno al 99% le fake news(6,7),e talora anche gli obiettivi delle organizzazioni occulte che progettano simili operazioni…

Nella prossima pillola : Le foto e le immagini Fake, la Topologia e Velocità di Propagazione delle "fake news" e la Guida finale per i comuni, indifesi utenti del Web…

Riccardo De Gobbi e Giampaolo Collecchia

Bibliografia

1) Ekman P. Friesen W: Constants across cultures in the face and emotion. J Pers Soc Psychol. 1971 Feb;17(2):124-9

2)Tiffany Watt Smith: The book of human emotion Profile Books UK 2015 https://profilebooks.com/

3)Ameeta Agrawal, Aijun An et Al.: Learning Emotion-enriched Word Representations http://creativecommons.org/licenses/by/4.0

4) Ameeta Agrawal, Aijun An: Unsupervised Emotion Detection from Text using Semantic and Syntactic Relations

5) A Beginner's Guide to Word2Vec and Neural Word Embeddings. https://wiki.pathmind.com/word2vec

6)Tacchini E,Ballarin G,de Alfaro Luca: Some Like it Hoax.Automated Fake News Detection in Social Networks. https://www.researchgate.net/publication/316471370_Some_Like_it_Hoax_Automated_Fake_News_Detection_in_Social_Networks/link/5a43af9ea6fdcce197189db3/download

7) Picazo Almira Josè: Le fake news e la loro formulazione matematica RBA Edit. Milano 2020

Perapprofondire:
Collecchia G. De Gobbi R.: Intelligenza Artificiale e Medicina Digitale. Una guida critica. Il Pensiero Scientifico Ed. Roma 2020
pensiero.it/catalogo/libri/pubblico/intelligenza-artificiale-e-medicina-digitale