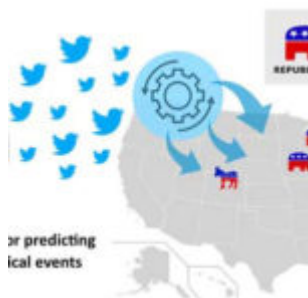


Chi vincerà le elezioni? Twitter (e l'intelligenza artificiale) aiutano a prevederlo



Si parte da un gruppo ristretto e ben selezionato di hashtag e se ne ricava, con l'aiuto dell'intelligenza artificiale, una previsione molto accurata delle intenzioni di voto degli elettori.

Il sistema, messo a punto dal gruppo di ricerca del professor **Domenico Talia**, si basa su reti neurali ed è già stato testato con successo durante le elezioni Usa del 2016 e le Politiche italiane del 2018. La metodologia – chiamata **IOM-NN** (Iterative Opinion Mining using Neural Networks) – si è dimostrata più affidabile dei tradizionali sondaggi e di altre tecniche di analisi social che sono basate sull'esame testuale dei messaggi, sulle emoji o sullo studio dell'andamento degli hashtag nel tempo e non usano reti neurali.

«All'inizio avevo qualche perplessità. Poteva bastare l'analisi di pochi termini per fare meglio dell'interpretazione di un intero messaggio? I risultati sono stati sorprendenti» commenta Domenico Talia.

L'intelligenza artificiale, che mostra ancora difficoltà nella comprensione dei testi soprattutto quando è necessario andare oltre il senso letterale, si comporta quindi molto bene quando

le viene chiesto di interpretare e prevedere il *sentiment* di una comunità partendo da parole chiave esplicite.

LA TECNICA – I ricercatori hanno selezionato una decina di hashtag, in base ai quali hanno estratto i messaggi da affidare all'analisi della rete neurale. Per gli Stati Uniti sono stati presi in considerazione 818.403 tweet pubblicati, tra il 10 ottobre e il 7 novembre 2016, da 141.959 utenti, mentre per l'Italia l'esame è stato condotto su 60.782 messaggi postati da 21.883 utenti tra il primo febbraio e il 3 marzo 2018.

In entrambi i casi il campione è stato individuato non solo in base agli hashtag, ma applicando una serie di filtri capaci di garantire l'autenticità degli account: la lingua in cui erano scritti i tweet e la localizzazione dell'utente nel Paese chiamato al voto, sulla base della georeferenziazione – dove disponibile – o delle dichiarazioni nella bio del profilo. Con l'utilizzo di tecniche già disponibili sono stati anche individuati e rimossi i bot.

A differenza di altri sistemi di analisi, la tecnica utilizzata dal gruppo di ricerca dell'**Unical** ha fatto attenzione poi anche alla prolificità del singolo utente. «Chi pubblica tantissimi post esprime comunque un solo voto – dice Talia – Abbiamo ripesato gli utenti molto attivi, per evitare che la quantità di tweet generati dallo stesso profilo falsasse il risultato». Il campione è stato infine bilanciato sulla base dell'età, perché fosse adeguatamente rappresentativo.

Ottenuto il set di dati, l'algoritmo ha iniziato a classificare i tweet, in base all'orientamento politico espresso. Non è stato sufficiente, per tutti, un solo tentativo: alcuni hashtag hanno costretto il sistema a iterare il processo di analisi. «Provando più volte, l'algoritmo impara da solo – spiega Talia – e nel nuovo tentativo riesce a

classificare quei tweet lasciati in sospeso prima». Nel caso degli Usa, ad esempio, il sistema in prima battuta non è riuscito a classificare i messaggi con gli hashtag *sex* e *woman*. Al secondo tentativo, imparando dai tweet in cui entrambi ricorrevano insieme ad hashtag pro Clinton, è riuscito a interpretarli correttamente.

Al termine, l'algoritmo è riuscito a classificare per l'Italia 24mila messaggi twittati da circa 10mila utenti, mentre per gli Stati Uniti ha assegnato circa 720mila tweet provenienti da 126mila utenti. Nel caso degli Usa, i tweet provenivano da 10 *swing states*, ovvero Stati in bilico nelle elezioni 2016: Colorado, Florida, Iowa, Michigan, Ohio, New Hampshire, North Carolina, Pennsylvania, Virginia e Wisconsin.

RISULTATI – La tecnica usata dai ricercatori Unical ha restituito una previsione di voto molto vicina ai risultati reali e ancora più accurata rispetto ai sondaggi tradizionali e ad altre tecniche di rilevazione basate su intelligenza artificiale.

Nel caso degli Usa, il sistema Unical ha correttamente previsto i risultati finali di 8 Stati su 10: i sondaggi si erano fermati a 6 e nessuno degli altri algoritmi presi in considerazione ha fatto meglio. Anche per le Politiche italiane del 2018 l'esito finale si è avvicinato molto alle percentuali reali ottenuti dai primi 4 partiti (M5S, PD, Lega e Forza Italia), con un grado di affidabilità maggiore rispetto ad altri approcci.

La ricerca è stata pubblicata su IEEE Access. Insieme al professor Domenico Talia ne sono autori anche i ricercatori del DIMES (Dipartimento di Ingegneria Informatica, Modellistica, Elettronica e Sistemistica) Loris Belcastro, Riccardo Cantini, Fabrizio Marozzo e Paolo Trunfio.

Link alla pubblicazione:

LEARNING POLITICAL POLARIZATION ON SOCIAL MEDIA USING NEURAL

NETWORKS