

I profili caratteriali di Twitter

Una metodologia che deriva tratti del carattere dagli aggiornamenti su Twitter è in fase di collaudo per aiutare le promozioni mirate e personalizzare i servizi clienti.

Tom Simonite

La comprensione dei sogni e dei bisogni di una persona – consapevoli e inconsapevoli – dal suo traffico e dai suoi acquisti on-line è divenuto un fattore cruciale per ogni genere di azienda. Ora, IBM si sta spingendo oltre con questa idea e sta collaudando una metodologia che presume i tratti psicologici principali di una persona analizzandone i post su Twitter. L'intento è quello di offrire un servizio personalizzato ai clienti e messaggi promozionali mirati.

«Dobbiamo andare oltre l'analisi comportamentale utilizzata per esempio da Amazon», dice Michelle Zhou, capo dello User Systems and Experience Research Group presso l'Almaden Research Center di IBM, in California, che ha sviluppato il software. «Vogliamo utilizzare i social media per derivare informazioni su una persona: qual è il suo interesse principale? quanto è dinamica la sua emotività?».

Il problema è se persone con personalità differenti abbiano bisogni differenti. Il software di Zhou sviluppa un profilo personale sulla base delle centinaia o migliaia di aggiornamenti recenti su Twitter. Questo profilo delinea "i cinque tratti principali" utilizzati comunemente nella ricerca psicologica: estroversione, gradevolezza, coscienziosità, nevrosi, apertura a esperienze. Delinea anche la misura di "valori" (per esempio, edonismo e conservatorismo) e "bisogni" (per esempio, curiosità e armonia sociale) di una persona.

Zhou dichiara di lavorare con diversi clienti IBM per verificare come la tecnologia potrebbe aiutare le loro imprese. Non dice di quali aziende si tratti, ma descrive come potrebbero utilizzare il sistema, per esempio, per regolare i messaggi promozionali inviati a e-mail o social media, o per selezionare il contenu-

to promozionale visualizzato quando un cliente accede al proprio profilo.

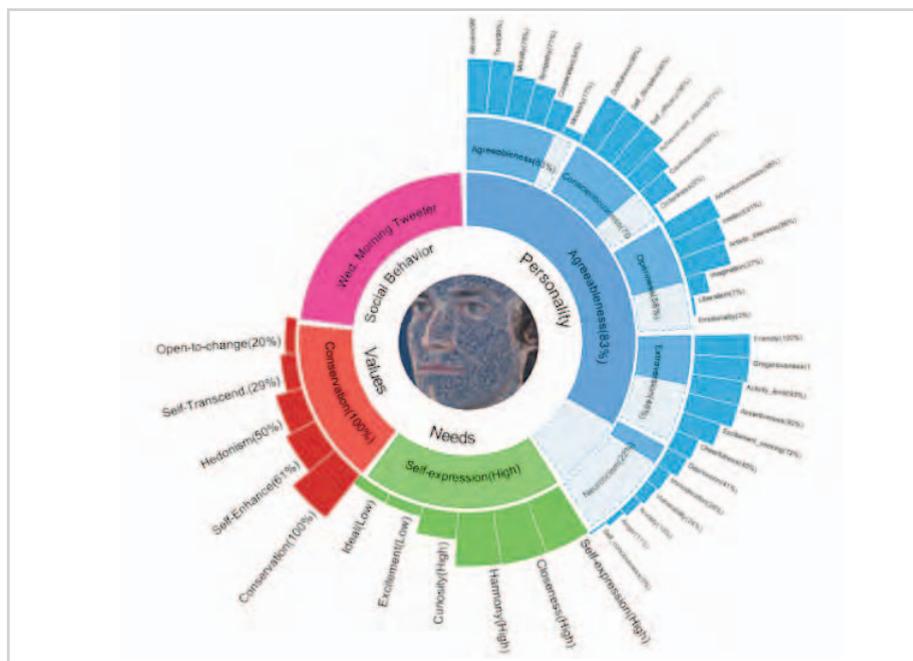
Una idea approssimativa della personalità di una persona potrebbe aiutare anche call center e altri servizi di assistenza ai clienti, come per esempio quando una Compagnia aerea deve informare i propri clienti che il loro volo è in ritardo o è stato cancellato. «Gli studi mostrano che le persone estroverse desiderano una ricompensa e un riconoscimento, come i punti per i viaggiatori frequenti», aggiunge Zhou. «Persone più critiche pretendono efficienza e vogliono ricevere immediatamente un aggiornamento sullo stato del loro volo». Nello scenario di un call center, il profilo caratteriale di una persona potrebbe suggerire a un operatore se fornire al cliente i "soli fatti" o assumere un approccio più personale e coinvolgente.

Molte imprese fanno già uso di software che analizzano le attività nei social-media e che però sono generalmente concepiti per aiutare i rappresentanti a interagire con clienti o riassumere il volume e il tono di una discussione, non per dedurre comportamenti significativi.

Stando al software analitico di IBM, l'account Twitter di Tom Simonite rivelerebbe una sua maggiore propensione a essere introverso che estroverso.

Il software IBM è stato sviluppato reclutando persone, sottoponendole a questionari psicologici e correlandone i risultati con la loro attività su Twitter, sulla base delle parole utilizzate. Queste correlazioni venivano quindi utilizzate per derivare modelli capaci di creare un profilo psicologico di una persona partendo unicamente dai suoi tweet.

In uno studio incrociato su 300 persone, la comparazione dell'analisi condotta sui loro profili Twitter con il loro profilo psicologico ha dato risultati "altamente correlati" per oltre l'80 per cento delle volte. Ciononostante, Zhou precisa che quando le persone utilizzano Twitter in maniera specializzata, per esempio da parte di giornalisti, i loro profili derivati dai tweet potrebbero non corrispondere. Ma Zhou sostiene che, siccome metodi attualmente in uso da parte delle aziende per identificare e comprendere i loro clienti sono relativamente imprecisi, il software IBM non deve necessariamente svelare per intero la personalità di una persona per rivelarsi utile. Inoltre, dovrebbe essere possibile adattare il software ad altre fonti di dati, quali per esempio le trascrizioni dei data center o i servizi chat per l'assistenza ai clienti. ■



I sentimenti di Twitter

La *sentiment analysis* sul Web dipende da come lo stato mentale di una persona venga espresso a parole. Un nuovo database permette di analizzarlo.

The Physics arXiv Blog

Una delle attività più in voga collegate ai siti sociali del Web è la *sentiment analysis*, ovvero la capacità di determinare le opinioni di una persona o la sua condizione mentale analizzando le parole postate su Twitter, Facebook o altri media.

Questo metodo ha inaugurato la possibilità di misurare la soddisfazione per i politici, per i film e per i prodotti, la capacità di gestire le relazioni con i clienti, di creare dialoghi per i giochi identitari, di misurare i flussi emotivi nei romanzi e così via.

L'idea è quella di automatizzare completamente questo processo, analizzando il flusso di parole prodotte nel Web grazie a tecniche avanzate di *data mining* (estrazione di dati) per valutare le opinioni su larga scala.

Tutto ciò dipende da quanto si è in grado di capire i sentimenti e le polarità (negativa o positiva) che le persone associano a ogni parola o combinazioni di parole. All'inizio di settembre, Saif Mohammad e Peter Turney del National Research Council Canada di Ottawa hanno reso pubblico un vasto database formato da parole e da sentimenti e polarità a esse associate, che hanno riunito velocemente e in modo poco dispendioso utilizzando il sito *crowdsourcing* Mechanical Turk di Amazon.

I due autori spiegano che questa procedura rende possibile aumentare la portata e la qualità del database in modo facile e veloce.

Molti psicologi credono che siano identificabili sei sentimenti fondamentali: gioia, tristezza, rabbia, paura, disgusto e sorpresa, o al massimo otto, se si includono fiducia e aspettativa: il compito di ogni glossario di parole riguardante i sen-

timenti determina quanto una parola sia strettamente associata ad ognuno di questi. Per farlo si utilizza un piccolo gruppo di esperti che già nel 1960 hanno creato un famoso database, conosciuto come General Inquirer, che contiene più di 11mila parole classificate con 182 differenti marche, che comprendono alcuni dei sentimenti che gli psicologi ritengono essere i più comuni.

Un database più attuale è il glossario *WordNet Affect*, che contiene alcune centinaia di parole marcate in questo modo da un piccolo gruppo di esperti che le hanno associate ai sentimenti più frequenti.

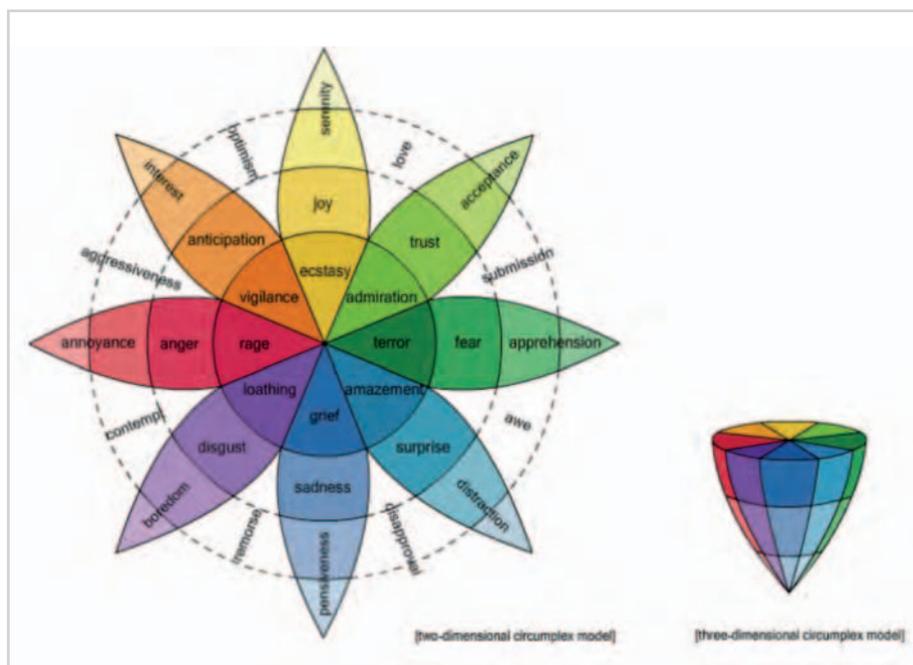
Uno dei problemi che si presentano con questo genere di strategia è il tempo richiesto dalla compilazione di un grande database. Mohammad e Turney hanno perciò provato un diverso approccio, selezionando circa 10mila parole da glossari preesistenti e associando a queste parole cinque caratteristiche emotive, che sono state poi sottoposte alla valutazione di oltre 2mila persone sul sito Mechanical Turk di Amazon, pagando 4 centesimi per ogni serie di risposte corrette.

Il risultato è stato un glossario contenente 10mila parole indicanti sentimenti,

combinare anche in locuzioni di due parole, che viene chiamato EmoLex. Un fattore importante in questa ricerca è la qualità delle risposte fornite dal *crowdsourcing*, tenuto conto che alcuni degli interpellati potrebbero rispondere a caso o inserire deliberatamente delle risposte sbagliate. Mohammad e Turney hanno affrontato questa possibilità inserendo nel test delle domande di controllo.

I due ricercatori hanno esaminato la qualità del loro database confrontandolo con quelli creati precedentemente da altri esperti e sostengono che regge bene il confronto: «Abbiamo confrontato una sezione del nostro glossario con i migliori dati esistenti e abbiamo potuto dimostrare che i nostri riscontri sono certamente di alta qualità».

Questo approccio presenta un'importante potenziale applicativo. Mohammad e Turney spiegano che dovrebbe risultare facile incrementare la dimensione del database e che la stessa tecnica può venire facilmente adottata per creare un analogo dizionario anche in altre lingue. Oltretutto lo si può fare in modo molto economico, se si pensa che i due autori in questo lavoro hanno speso solo 2.100 dollari. ■



Fuoco alle corde!

La rabbia si diffonde più velocemente e ampiamente della gioia, sostengono gli informatici che hanno analizzato le dinamiche dei sentimenti su Weibo, il Twitter cinese.

The Physics arXiv Blog

Una caratteristica ben conosciuta dei social network è che persone simili tendono ad attrarsi a vicenda: ogni simile ama il suo simile. Perciò una domanda interessante è se queste somiglianze fanno sì che le persone si comportino nello stesso modo quando sono on-line, se cioè possono indurre un comportamento di massa.

Oggi, abbiamo un'interessante rappresentazione di questo fenomeno grazie al lavoro di Rui Fan e dei suoi colleghi della Beihang University in Cina. Questi ricercatori hanno confrontato il modo in cui i tweet caratterizzati emotivamente influenzano le altre persone in rete. La loro conclusione è sorprendente. I risultati dimostrerebbero chiaramente che la rabbia ha più influenza rispetto ad altri sentimenti come la gioia o la tristezza; una scoperta che potrebbe avere delle implicazioni significative per la comprensione del modo in cui le

informazioni si diffondano attraverso la rete sociale. Questi ricercatori hanno avuto i loro dati da Weibo, un servizio simile a Twitter che è diventato incredibilmente popolare in Cina. In soli quattro anni ha attirato più di 500 milioni di utenti che postano 100 milioni di messaggi ogni giorno.

Nel 2010, per sei mesi, Rui ha raccolto con i suoi colleghi quasi 70 milioni di tweet da 200 mila utenti e costruito un social network in cui gli utilizzatori interagiscono gli uni con gli altri, per esempio inviandosi messaggi o inoltrandosi i tweet a vicenda.

Rui e colleghi, per assicurarsi di avere studiato solo persone solidamente collegate, hanno incluso nel lavoro quanti avevano più di 30 interazioni durante quel periodo. Successivamente, hanno determinato il sentimento di ogni tweet nel loro database analizzando gli emoticon che contenevano e li hanno divisi in quattro categorie espressioni gioia, tristezza, rabbia o disgusto. Infine hanno studiato come i sentimenti si diffondano attraverso la rete. Per esempio, quando una persona spediva un tweet che esprimeva rabbia, era perché il destinatario aveva inviato a sua volta un messaggio di rabbia e probabilmente la catena di connessioni sarebbe così proseguita.

I risultati furono sorprendenti. Quando si è trattato di tristezza e disgusto, Rui e i suoi colleghi hanno trovato poche correlazioni tra gli utenti. Tristezza e disgusto non si diffondono facilmente attraverso la rete, mentre hanno trovato una elevata correlazione tra gli utilizzatori che hanno twittato messaggi di gioia.

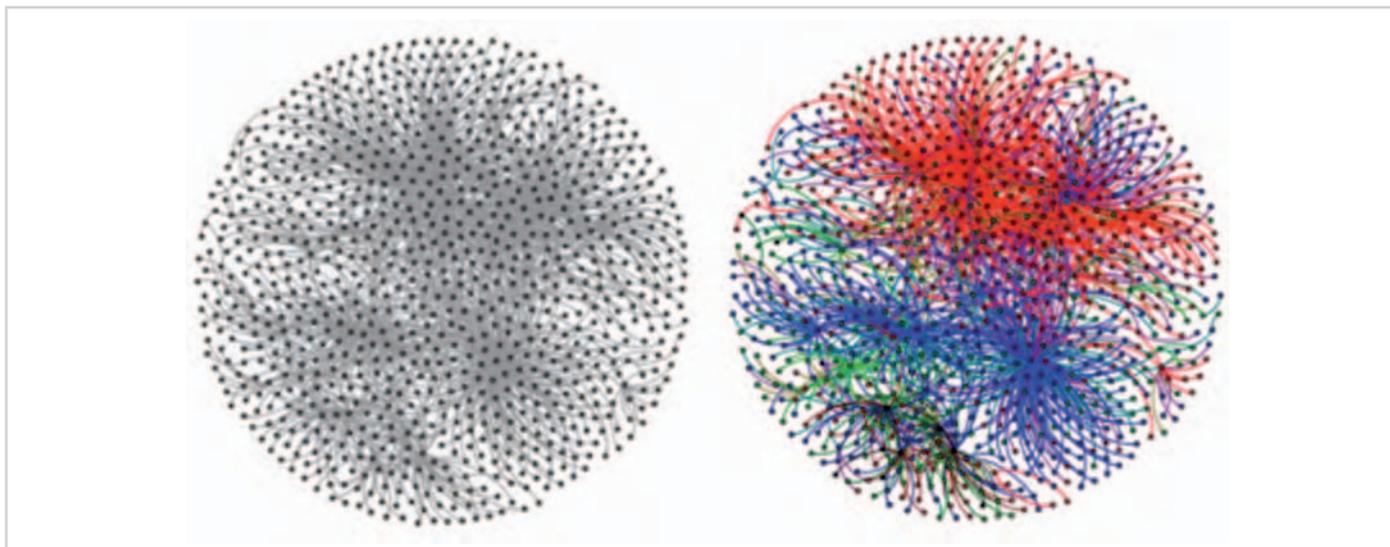
Ma la più alta correlazione era tra gli utenti che esprimevano rabbia. Rui e colleghi dicono che la rabbia influenza fortemente il luogo in cui appare, diffondendosi in media per almeno tre passaggi. «La rabbia ha una maggiore forza di correlazione rispetto agli altri sentimenti» e quindi, probabilmente, si diffonde più velocemente e ampiamente in rete. I ricercatori che hanno studiato il contenuto di molti tweet di rabbia, sostengono che due generi di eventi sembrano scatenare questo tipo di messaggi.

I primi sono i conflitti tra la Cina e i paesi stranieri, come le attività militari degli Stati Uniti e della Corea del Sud nel Mar Giallo e la collisione tra una nave cinese e una giapponese avvenuta nel 2010.

I secondi sono i problemi sociali interni come la sicurezza alimentare, la corruzione del governo e la demolizione di case per i nuovi insediamenti, tutti argomenti caldi in Cina. «Tutto ciò può spiegare perché gli eventi collegati a problemi sociali si propagano in modo estremamente veloce su Weibo».

Naturalmente sarebbe interessante vedere se lo stesso effetto possa osservarsi nei network occidentali come Twitter, ma dovrebbe risultare relativamente semplice scoprirlo, dato il crescente interesse per la *sentiment analysis* e gli strumenti di ricerca a disposizione.

La conclusione è che quando si tratta di diffusione dell'informazione, la rabbia è molto più potente delle altre emozioni. Così se vuoi diffondere il tuo messaggio, lascia pure che il sentimento che nasconde, possa venire fuori. ■



Come ti muovi...

Gli scienziati informatici spiegano come l'aumento nel volume di tweet localizzati geograficamente stia rivelando le variazioni degli spostamenti internazionali tra paesi e culture.

The Physics arXiv Blog

Uno degli interessanti effetti secondari del fenomeno Twitter è la prospettiva totalmente nuova che offre allo studio della popolazione umana. In particolare, un numero crescente di tweet contiene informazioni sulla localizzazione di chi sta twittando. Diversi gruppi stanno studiando la distribuzione spaziale dei tweet e come questi varino da un luogo a un altro.

Bartosz Hawelka dell'University of Salzburg, in Austria, insieme ad alcuni colleghi, ha spinto oltre questa idea, analizzando la geolocalizzazione dei dati per studiare eventuali schemi nascosti negli spostamenti internazionali. Stando alla loro ricerca, i messaggi geolocalizzati di Twitter rivelerebbero un modello globale sulla mobilità umana mai studiato prima e i dati mostrerebbero come questi schemi di viaggio varino tra le culture, i paesi e anche nel corso dell'anno.

I ricercatori hanno iniziato con un gruppo di dati di circa un miliardo di tweet, geolocalizzati tra gennaio e dicembre 2012. Hanno quindi rimosso tutte le attività di Twitter che non riguardavano attività umane, come per esempio le pubblicità e alcuni tipi di giochi. Hanno anche escluso ogni account di Twitter in cui i dati dell'utente implicavano un trasferimento a una velocità superiore a quella di un aereo passeggeri. Hanno infine determinato la residenza di ogni utente analizzando le posizioni dei luoghi da cui twittava, indicando come luogo di residenza quello che presentava la maggiore attività su Twitter.

Questi risultati hanno prodotto un database dei "residenti" di Twitter in 243 dei 253 paesi del mondo. Gli Stati Uniti sono la nazione che utilizza maggiormente Twitter, con più di 3,8 milioni di utenti, seguiti da

Regno Unito, Indonesia, Brasile, Giappone e Spagna, tutti con più di 500mila utenti.

Per avere una idea del tasso di penetrazione in ogni paese, i ricercatori hanno anche calcolato il rapporto tra il numero di utenti di Twitter e la popolazione complessiva. Questo tasso varia di paese in paese e cresce in maniera scalare con il PIL *pro capite* di ogni paese, per cui i paesi più ricchi hanno più utenti di Twitter *pro capite*.

Successivamente, il gruppo di Hawelka ha calcolato in ogni paese la percentuale di utenti che nel corso del 2012 hanno visitato un altro paese. Ne è emerso che i paesi con gli utenti più mobili sono Hong Kong, Belgio e Austria. Ciò non deve sorprendere, perché questi sono territori piccoli con un facile accesso verso altri paesi. Più sorprendente è la scoperta che il paese con il minore numero di utenti che viaggiano sono gli Stati Uniti, dove i tweeter sono meno mobili rispetto a quelli di Indonesia, Venezuela, Arabia Saudita e di ogni altro paese.

Il gruppo di Hawelka ha calcolato inoltre la distanza media percorsa dai tweeter nei differenti luoghi, una misura che ha chiamato raggio della spirale (o raggio di rotazione). I paesi in testa tendono a essere i più isolati. I tweeter in Australia e Nuova Zelanda, per esempio, hanno un raggio di rotazione medio di 700 chilometri. Non è una distanza enorme se considerata su scala globale: «Dimostra che anche nell'era della globalizzazione le persone ancora tendono a viaggiare localmente, visitando paesi confinanti con maggiore assiduità rispetto a quelli più distanti».

Infine, Hawelka e i suoi collaboratori hanno studiato come gli schemi dei viaggi cambino nel corso del tempo. Sostengono che, generalmente, i tweeter tendono a essere più mobili nel fine settimana e più disposti a viaggiare all'estero durante i mesi estivi di luglio e agosto e anche alla fine dell'anno. I modelli più interessanti emergono quando si esaminano i modelli temporali di viaggio nei singoli paesi o in gruppi particolari. I paesi arabi, per esempio, mostrano un modello di spostamento che ha un chiaro significato culturale: «Il periodo del Ramadan corrisponde a una forte riduzione nei viaggi all'estero, mentre il periodo del pellegrinaggio alla Mecca segna un forte picco alla fine di ottobre».

Ma quanto questi dati sono rappresentativi? Un problema potrebbe derivare dalla

possibilità che i tweeter creino gruppi speciali all'interno di ogni società, perché tendono a essere appassionati di tecnologia. Queste tendenze rappresentano la popolazione in generale o solo una piccola parte?

Il solo modo per verificarlo è confrontare i dati di Twitter con i tradizionali sistemi di misurazione degli schemi di viaggio internazionale. Anche se informazioni così dettagliate sono difficili da raccogliere, in altri modi, su scala globale, il gruppo dei ricercatori ritiene che esistano strette correlazioni tra i loro risultati e le statistiche del turismo globale. Se ciò dovesse rivelarsi corretto, la loro ricerca fornirebbe un nuovo e importante strumento per studiare una delle caratteristiche della vita moderna: il cosiddetto viaggio su scala globale.

Anche i dati raccolti dai cellulari permettono una visione precisa dello spostamento, con una risoluzione limitata solo dalla distanza tra le stazioni di trasmissione. Diversi gruppi di ricerca hanno fatto buon uso di questi dati. I dati dal cellulare, però, raramente superano i confini internazionali e certamente non hanno la portata che offre Twitter.

Hawelka progetta ora di utilizzare i dati di Twitter per esaminare gli spostamenti su una scala molto più precisa.

A un numero crescente di tweet sono associate coordinate GPS che danno la posizione dell'utente all'interno di un'area di alcuni metri: «La prossima ricerca cercherà di comprendere fino a che punto queste conoscenze possano tradursi in scale spaziali più raffinate». ■

