Quest'anno vincono i dati

Dopo la lezione americana di Obama, anche la politica di casa nostra studia i social network. Ecco chi sono gli italiani che sanno interpretare il vero sentiment della rete. Gli abbiamo chiesto: chi ha davvero i numero per vincere?

Stefano Iacus è uno statistico. Ha 44 anni e uno dei suoi sogni è cambiare la politica.

Non è però l'ennesimo volto della società civile che ha deciso di (provare a) entrare in Parlamento: lui la politica la vuole cambiare da fuori. Grazie alla matematica di una formula (<<E molto lavoro>>) Stefano potrebbe essere infatti determinante alle prossime elezioni politiche del 24 e 25 febbraio. Quella che avete appena letto è la base del lavoro di Iacus. Per capire come questa formula possa essere decisiva per la politica Italiana (al pari dello Spread o della riduzione delle tasse) è necessario tornare indietro nel tempo, alla fine del 2010, e volare in Giappone. Stefano, ricercatore dell'università Statale di Milano, è a Tokyo per lavoro. Per caso, in quegli stessi giorni, in città c'è un altro protagonista di questa stessa storia: il politico quarantenne Luigi Curini, anche lui della statale grande e anche lui della capitale giapponese perché collabora con l'università di Tokyo. I due si conoscono ma non hanno mai lavorato insieme. Così, una sera, si ritrovano in un bar a scambiare quattro chiacchiere sui progetti a cui vorrebbero dedicarsi nei mesi successivi. Scoprono così che stanno pensando alla stessa cosa. L'idea è usare la grande mole di dati che si trova su blog e social media (soprattutto FACEBOOK e TWITTER) per condurre analisi del sentiment (cioè dell'umore delle persone) su qualsiasi tema. Insomma, mettere a punto un sistema che riesca a interpretare e a sintetizzare tutto quello che si dice in rete.

Tornati in Italia, Stefano e Luigi si mettono in contatto con un altro collega: l'economista Giuseppe Porro, 49 anni. Bisognava costituire un gruppo di lavoro per portare avanti il

progetto che avevano in mente. “Ormai erano diversi gli stufi scientifici che dimostravano come fosse possibile analizzare il sentiment su un determinato argomento a partire dai social media”, mi racconta Luigi mentre siamo seduti in una delle stanze dell’università dove è nato Voices from the Blog (VfB), il progetto di ricerca che i tre accademici hanno messo in piedi nel 2011.

Nei primi mesi c’è la fase preparatoria. “Avevamo deciso di puntare su un algoritmo realizzato da due scienziati statunitensi dell’università di Harvard, Daniel J. Hopkins e Gary King”, spiega Stefano. L’equazione è questa: $P(S) = P(S|D)$. “Dovevamo solo adattare alle nostre esigenze questa formula, sulla quale i due ricercatori americani hanno fondato una società di consulenza aziendale che analizza il gradimento online di prodotti e servizi”, prosegue lo statistico.

L’obbiettivo del team milanese era quello di cimentarsi pria di tutto con la politica. E il primo banco di prova arriva nella primavera del 2012 con le elezioni presidenziali francesi. Al primo turno le cose vanno abbastanza bene, ma non perfettamente. Per Sarkozy la differenza tra il dato reale e quello previsto è dell’1%, mentre per Hollande è del 5%, in meno rispetto ai risultati delle urne. Molto meglio al secondo turno: la vittoria del presidente socialista viene giustamente prevista, anche se con uno scarto di 3 punti a favore di Hollande. In un report lo stesso staff di VFB leggo che questo errore sarebbe stato causato dal sottovalutazione del tasso di astensione.

I ricercatori si rimboccano le maniche per “aggiustare” l’algoritmo e si preparano al successivo grande appuntamento elettorale: le presidenziali americane del 6 novembre 2012. E’ un evento importante per il lavoro della squadra di Iacus. Per la prima volta si confronta con i matematici d’oltreoceano che, per i candidati o da indipendente, hanno seguito la campagna elettorale. Barack Obama e Mitt Romney, infatti, hanno nei loro staff ingegneri, matematici e statistici che studiano ogni dato (provenire dalla rete e non solo) per definire scientificamente le mosse politiche durante la

campagna.
Dall'ideale duella a distanza VfB ne esce bene. Grazie all'analisi di diversi milioni di tweet, il team milanese predice esattamente i risultati degli stati chiave, quelli in bilico, dove Romney e Obama si giocavano la partita. E' sorprendente anche il risultato delle percentuali generali: la differenza reale tra i due candidati è del 2,2%, mentre quella pronosticata da VfB è appena superiore: 3,5%. La stessa performance si registra qualche settimana più tardi con le primarie del centrosinistra in Italia: l'errore medio nei risultati dei cinque candidati è di 1,96%.
"Di solito l'errore medio con la formula di Hopkins e King è del 2%, con gli ultimi lavori siamo riusciti a tenerci sotto", commenta Iacus mentre Curini mi spiega nel dettaglio come vengono fatte le rilevazione: "Inizialmente analizzavamo solo i blog su un determinato tema: scaricavamo tutti i post e iniziava l'indagine. Ora ci siamo resi conto che lo strumento migliore è Twitter: ci sono molte più voci e, soprattutto, li trovi le reazioni immediate, fondamentali per capire in tempi brevi l'umore delle persone. Ci agevola anche la crescita costante del numero di utenti che sua Twitter dallo smartphone".
Grazie alle Api di Twitter, lo staf di Vfb riesce a scaricare in maniera automatizzata tutti i tweet relativi a un argomento. Successivamente si mette al lavoro una squadra di persone in carne e ossa che fa un'analisi semantica di un campione dei contenuti scaricati. "Per risultare affidabile il campione deve essere di un migliaio di tweet", dice Iacus. "A questo punto i altri esperti di codifica dividono tutti i post in categorie e a ogni categoria associano delle parole chiave". Successivamente estendono questa categorizzazione a tutto il resto dei tweet scaricati. Ed è qui che entra in gioco la formula $P(S|D)^{-1} \times P(S) = P(D)$.
In questa equazione il risultato P(D) – cioè la distribuzione delle preferenza in un insieme (nel nostro caso di tweet) – deriva dalla moltiplicazione tra il reciproco di P (S|D) – cioè la probabilità che venga utilizzata una certa successione

di parole tenuto conti che si vuole esprimente un determinato contenuto semantico, un opzione – e P(S), ovvero la frequenza delle parole delle persone stesse. Così, estendendo questa formula dal campione al resto del database si ottiene il risultato complessivo, che può comprendere l'analisi di milioni di tweet.
Il sistema di VfB non è completamente automatizzato, ma passa per un esame "umano" dei contenuti, perché secondo i ricercatori milanesi i dizionari ontologici informatici non sempre dosi o in grado di interpretare correttamente un testo. Per spiegare meglio il concetto fanno un esempio: la frase "una bella fregatura" contiene al suo interno una parola solitamente positiva (bella) e una negativa (fregatura), me è chiaro che il suo senso complessivo sia da considerarsi esclusivamente negativo.
Lavorando così, quindi, il tempo di VfB è riuscito a prevedere l'esito di tutte le ultime elezioni importanti, sia all'estero sia in Italia. Non mi stupisce affatto che l'impresa politico degli elettori statunitensi possa emergere anche da Twitter, ma mi domando come questo sia possibile nel nostro paese, dove poco più della metà degli italiani si collega abitualmente alla rete e (secondo i dati Audiweb/Neielsen di settembre 2012) ci sono appena 3,64 milioni di utenti iscritti alla piattaforma microblogging. Mi risponde Iacus, che è anche tra gli sviluppatori del linguaggio di programmazione R sul quale gira il tool open source con cui VfB esegue tutti i suoi calcoli: "Un tempo la distanza tra chi frequentava o meno la rete era molto grande. Oggi non è più così. In questi mesi stiamo proprio lavorando a una ricerca che dimostri come le opinioni che si esprimono online si formino fuori da internet e quindi chi è sul web è solo un portavoce di quello che avviene in tutta la società, sia essa analogica o digitale. In questo modo si spiega come mai le nostre analisi, che si basano su Twitter, rispecchino gli umori e il giudizio che si sono anche offline".
I risultati di Iacus e del suo team hanno colpito molti. Per questo (anche grazie a una collaborazione con il sito del

corriere della sera che gli ha dato visibilità) diversi politici, di tutti gli schieramenti, sono entrati in contatto con VfB e gli stanno commissionando analisi per capire come muoversi durante la campagna elettorale. “Le richieste, però, vanno anche oltre la politica”, sottolinea Curini. “In questi mesi sono arrivate da noi anche banche e aziende che operano nei settori della cosmetica e del beverage”. Ecco perché il gruppo VfB (che nel corso degli anni si è arricchito anche nel politologo 29 Andrea Ceron) ha deciso di trasformarsi in un’azienda. “Abbiamo costituito una Srl, uno spinotto dell’università di Milano che mantiene una piccola quota azionaria. Per qualche anni ci sarà consentito di fare sia gli accademici sia gli imprenditori, poi dovremmo decidere se diventare una società completamente indipendente oppure uscire dall’azienda”, spiegano.

L’offerta commerciale di VfB è ancora agli inizi. Una proposta più complessa è invece in grado di farla BlogMeter, una società nata nel 2006 con sedi a Milano, Torino e Roma. Inizialmente ha sviluppato in motore per le analisi semantica delle conversazioni italiane in internet che esplorava soprattutto blog, focus e newsgroup. Nell’ ultimo anno, però, i principali territori di osservazione sono ovviamente diventati social media, in particolare Facebook e Twitter. A mostra i quello che sono in grado di fare gli strumenti di BlogMeter è Vincenzo Cosenza, social media strategist e responsabile della sede romana, che in rete è diventato famoso per il suo osservatorio social media in Italia e Facebook e per la Mappa mondiale dei social network le cui edizioni vengono pubblicate dai principali giornali internazionali.

I dati e i grafici (con un interfaccia facilmente comprensibile) che mi fa vedere Cosenza riguardo i principali leader politico di questa campagna elettorale: “L’analisi su Facebook, per esempio, non solo mostra in tempo reale il numero di like alla pagina o al profilo del politico ma anche (e questo è uno dei dati più importanti, ndr) il numero di persone che interagiscono con una pagina o un singolo contenuto, per esempio status o foto”, spiega. La stessa cosa

avviene su Twitter e in entrambi i social media è possibile monitorare anche il tempo medio in cui un politico risponde alle sollecitazioni degli utenti. Oggi tutti questi dati vengono utilizzati insieme ai classici sondaggi elettorali ma per alcuni aspetti sono migliori. Strumenti come quelli di BlogMeter o VfB ascoltano quello che si dice in rete (quindi nella società) e il dato non è alterato dalla sollecitazioni esterne come la domanda di un sondaggista. Qualche mese fa, per esempio, nessuno avrebbe preparato un sondaggio elettorale in cui fosse prevista la possibilità di votare il Movimento 5 stelle, rischiando di sottovalutare un fenomeno fondamentale per cercare di prevedere l'esito delle votazioni.
Il team di lavoro di BlogMeter è un vero e concentrato di profili scientifici: "C'è chi si occupa di capire le funzionalità che devono avere i nostri crawler (i software che analizzano i contenuti di una rete, ndr) e intelligence, sviluppatori e sistemisti che realizzano questi tool e gestiscono gli aspetti tecnologici, linguistici computerazionali che insegnano alle macchine come interpretare i messaggi raccolti e identificarne i sentimenti, analisi che estraggono indicazioni quanti-qualitative utili dai dati raccolti", racconta Cosenza.
"Il nostro motore semantico" dice la linguista computerazionale di BlogMeter Vanessa Nardone, "segue gli sbalzi di umore attraverso tecnologie di interpretazione automatica del linguaggio e di classificazione delle opinioni espresse in rete, fa sentiment analyis. Il processo di comprensione automatica dell'opinioni si articola in tre fasi: l'esame di porzioni di testo e di indicatori testuali come la punteggiatura; l'estrazione del mood positivo e negativo dei messaggi mediante la previa analisi del testo; la classificazione dei documenti secondo una polarità positiva, negativa o mista. Al calcolo della polarità il motore High, medium, low; alto, medio o basso – che descrive l'intensità con cui l'opinione è espressa nel documento".
Questi strumenti vengono usati per esempio dall'Udc che ha preparato la campagna elettorale anche a partire dai dati che

gli forniscono BlogMeter. “Prima di iniziare la campagna abbiamo condotto un’analisi sulle conversazioni e le parole più rilevanti in rete e cos’ abbiamo tirato fuori gli argomenti di maggiore interesse”, racconta Marco Tosi, consulente Udc per la comunicazione sui media digitali. Grazie a questa indagine è stato possibile sia elaborare un programma che coinvolgesse almeno l’80% dell’elettorato italiano (punto su lavoro, Sud – perché Monti è troppo visto come un uomo del Nord – e donne) sia preparare i discorsi dei leader Udc, per esempio stilando una lista di possibili risposte nei talk show. Con i dati di BlogMeter lo staff del partito di Casini è riuscito anche ad affinare la strategia di comunicazione. Per esempio, si è deciso di non puntare su famiglia e pensioni nella campagna online , perché si è osservato che gli elettori interessati a questi temi non stanno su internet.
“In Italia si sta facendo qualcosa sulla scorta di quanto accade negli USA, dove la rete fondamentale nella vittoria alle elezione. Ma siamo ancora all’inizio”, dice Dino Amenduni, responsabile nuovi media e consulente di comunicazione politica di Proforma, l’agenzia che ha seguito le campagne elettorali del sindaco di Bari Michele Emiliano, nel 2009, e del governatore della Puglia Nichi Vendola, nel 2010: “In entrambi i casi abbiamo usato internet e un analisi delle conversazioni online per contribuire a stilare parte del programma. Ma c’è ancora bisogno di un salto culturale: i politici devono imparare a non considerare una sorta di lesa maestà il ricorso figure come quello dello spin doctor”.
Un salto che in molte aziende italiane è stato già fatto. Poste mobile è l’operatore del Gruppo Poste Italiane: è presente sul mercato dal 2007, ha oltre 2 milioni di clienti e usa il sistema di monitoraggio di BlogMeter. “Questo strumento ci permette di capire qual è il grado di soddisfazione dei nostri clienti, in quanto tempo siamo in grado di rispondere alle domande che ci arrivano sui social media e dove possiamo migliorare”, spiega Carlo Cortesi, responsabile new media e web di Poste Mobile. I report vengono discussi anche nelle riunioni in cui si stabiliscono le offerte commerciali; e il

pannello di controllo è accessibile ai team del marketing, del servizio clienti e della comunicazione. Non vengono comunque monitorate solo le discussioni su Poste Mobile, ma quelle sull'intero mercato della telefonia, concorrenza compresa.
Chi lavora con grandi brand è anche Tweetminster, la terza protagonista di questa storia che, a differenza di VfB e BlogMeter non opera in Italia, ma in Inghilterra. A fondarla nel 2008 a Londra è stato però il romano Alberto Nardelli (oggi 32 anni). "c'erano appena 4 parlamentari inglesi su Twitter – ora sono più di 400 – e anche i giornalisti erano pochissimi. Abbiamo però scommesso che la piattaforma sarebbe cresciuta e già a metà del 2009 i politici erano diventati 100 mentre i media incominciarono a interessarsi al fenomeno", ricorda Nardelli. "Per noi la svolta è arrivata nel 2010, quando ci sono state le elezioni politiche in Inghilterra (vinte dal conservatore David Cameron, ndr): abbiamo incominciato a realizzare delle mappe per vedere dove i candidati stavano facendo campagna elettorale, a monitorare i trend nelle discussioni politiche e anche a ricostruire i network di relazioni tra i politici".
In questo modo l'azienda di Nardelli ("Siamo in 5: io mi occupo di prodotto e strategia; gli altri sono designer e sviluppatori") si è fatta notare ed è stata assoldata da Reuters, BBC, diversi governi, ambasciate e Unione Europea per analizzare le conversazioni che nascono attorno a determinati temi o a decisioni politiche.
Il cuore della tecnologia – interamente sviluppata in casa – si basa su analisi di trend all'interno di comunità molto specifiche e selezionate (nel caso della politica: ci sono politici, giornalisti, analisti, Think Tank, apparati di partito). Guardando dentro questa comunità si scoprono il link più condivisi, le questioni più dibattute. Gli argomenti a cui gli esperti, i media e l'opinione pubblica danno più peso. "Il nostro punto di partenza sono quindi i Tweet dentro una comunità", racconta Nardelli "Costruiamo la rete, identifichiamo i trend, estraiamo i link per aggregarli e misurarne la popolarità. Poi selezioniamo gli individui che

sono più attivi e menzionati. In oltre abbiamo messo appunto un sistema che cerca di capire quando qualcosa sta succedendo e attiva alert. Lo proporremo ai nostri clienti nella prima parte di quest'anno come servizio a sé. Lo chiamiamo Pre Breking News: vuol dire catturare notizie prima che arrivino sulle agenzie".

Una grossa fetta dei clienti dell'agenzia britannica è rappresentata da società finanziarie per le quali avere una notizia anche pochi secondi prima può essere cruciale. " Già oggi Dow Jones paga 300mila dollari all'anno per un servizio che gli offre notizie esclusive 60 secondi prima che arrivino sui canali normali delle agenzie", dice Nardelli. "Noi vogliamo fare la stessa cosa per altri ambiti, per esempio quello politico".

Le analisi dei dati nei social media (e le scelte e le previsioni che non derivano) trovano infatti nella politica solo una parte delle proprie applicazioni. Ci sono già diversi studi scientifici che dimostrano come l'osservazione di Twitter e simili può anticipare gli andamenti in Borsa o aiutare a comprendere la grandezza di un'area colpita da un terremoto o la propagazione di un epidemia. Tutte le informazioni raccolte in rete servono per accorciare i tempi di reazione di una determinata situazione. Negli Usa, per esempio, le grani case di produzione hollywoodiane stanno usando una ricerca condotta agli Hp Lbs di Palo Alto, California, che dimostra come sia possibile anticipare gli andamenti del botteghino di un film partendo dai commenti che vengono espressi su Twitter. Insomma, le tecnologie come quelle di VfB, BlogMeter e Tweetminster verranno usate sempre di più e sempre più spesso per comprendere la società. E non c'è bisogno di un analisi sui social media per sapere che la probabilità che aziende come queste avranno successo è davvero molto alta.