

Google modererà i vostri commenti con un algoritmo intelligente

di Marco Vitali

Stavo curiosando su Facebook, “scrollando” come si dice, quando d’improvviso sono stato colpito da un avviso mai notato prima. Sono abbastanza sicuro che sia una novità, una caratteristica introdotta da poco. Eccolo qui:

La modalità selezionata è “Commenti più rilevanti”, pertanto alcune risposte potrebbero essere state filtrate.

Fra i commenti ad un post, ecco spuntare la frase qui sopra, per avvertirmi che per migliorare la mia esperienza un bot automatico ha moderato i commenti e filtrato automaticamente quelli meno rilevanti o non inerenti, lasciando solo quelli davvero interessanti. Poca cosa, direte voi. Niente di particolarmente interessante? Forse, ma l’utilizzo di un bot o di un algoritmo per moderare i commenti ad un post apre scenari decisamente interessanti.

Verrà un bot e ci seppellirà

La domanda che sorge spontanea è questa: può un algoritmo matematico, un cavillo informatico, calcolare automaticamente l’interesse che suscita un commento? Risposta semplice: no. Risposta complessa: no, ma ci stiamo provando.

Google ha recentemente valutato l’idea di utilizzare l’intelligenza artificiale al fine di moderare i commenti. Come? Attraverso una nuova mirabolante tecnologia che promette di riconoscere offese, violenze verbali, *flame*, commenti fuori tema e così via. Il *flame*, per chi non dovesse avere familiarità con la parola è un messaggio volutamente offensivo e provocatorio, tipico di chi cerca di far degenerare una conversazione. Una delle principali regole della netiquette è proprio “*non flammare*” (*don’t flame*), non fare *flame*. Le altre sono riassunte qui sotto.

Oggi il lavoro del “moderatore di commenti” è svolto da uomini in carne e ossa che ogni giorno valutano centinaia di commenti decidendo quali salvare e quali invece cestinare o nascondere. Possiamo sperare di utilizzare un bot automatico al loro posto? Il *software Perspective* promette di farlo.

Perspective è un *tool* che aspira a suddividere automaticamente i commenti degli utenti in “interessanti” e “non interessanti, ma anche in “offensivi” e “inoffensivi” o “spam” e “non spam”. Funziona attraverso un principio di apprendimento automatico: analizzando centinaia e migliaia di

commenti il software immagazzina esempi su esempi. Trovandosi poi di fronte situazioni simili capisce da solo cosa fare, fornendo anche possibili soluzioni (eliminazione dell'utente, del commento ecc.).

È possibile testare il programma [a questo link](#). Da quanto ho potuto vedere accetta solo frasi scritte in lingua inglese.

Qual è il colore più offensivo?

Ovviamente appena sbarcato sulla pagina linkata sopra ho tentato di mettere in difficoltà il programma con le frasi più disparate. Ciò che *Perspective* restituisce è una percentuale che indica quanto una certa parola, frase o espressione sia ritenuta offensiva o flammatoria per gli utenti. Fra le varie incongruenze e imperfezioni trovate, ne cito soltanto una. Il nero, “*black*”, è il colore più offensivo di tutti. Purtroppo è piuttosto facile capirne il perché.

Già questo ci mette di fronte a una serie di problemi che difficilmente la macchina potrà risolvere. Come può un calcolo matematico distinguere fra il nero “persona-di-colore” e il nero “colore-delle-cose”? La risposta è davvero di difficile soluzione, fatto sta che la percentuale che leggete sopra, il 49%, indica che una volta su due la parola “*black*” è associata a un commento offensivo, o non inerente, o non interessante. Anzi, per meglio dire, che la parola “*black*”, è somigliante al 50% ad un commento offensivo, non inerente o contenente spam.

È uno sporco lavoro, qualche androide deve pur farlo

Ogni secondo, su Facebook (e stiamo parlando solo di uno delle molte piattaforme disponibili), vengono caricati circa 41mila post. Ciascuno di questo otterrà quanti commenti in media? Che siano due, cinque o cinquanta, è evidente che non si potranno mai assumere abbastanza moderatori per garantire un servizio di qualità e discussioni sempre stimolanti e sotto controllo.

Ecco che gli androidi, i bot, diventano indispensabili. Anche se, evidentemente, non possono ancora maneggiare il linguaggio come un umano, comprendere i sensi e i significati (come direbbe [Gottlob Frege](#)), comprendere l'ironia e anche il nonsense e il paradosso.

Così come i mezzi di trasporto, l'industria manifatturiera e altre decine e decine di branche e attività dell'uomo, anche la comunicazione rivendica i suoi robot e i suoi automatismi. Forse non siamo ancora pronti, ma il passo ancora da compiere, c'è da scommetterci, è breve e alla nostra portata.