(2 Marzo 2002)

INFORMATICA

Algoritmo, dimmi chi l'ha scritto

di Alice Andreoli

Può un computer riconoscere l'autore di un testo? La risposta è sì. Risolvere, per esempio, l'annosa questione della paternità di alcuni testi, soprattutto antichi, è ora possibile. Ma anche datare un'opera letteraria in base allo stile di scrittura. Tutto grazie a un algoritmo, utilizzato da tre ricercatori dell'Università La Sapienza di Roma, che è in grado di stabilire in quale lingua un testo è scritto e anche di attribuirne l'autore.

Il metodo, riportato su Physical Review Letters del 28 gennaio scorso, si basa sulle stesse tecniche di zipping usate per comprimere i file senza perdere informazioni. Per zippare un testo non è necessario conoscere l'alfabeto, la grammatica o la sintassi, esso rappresenta, per un computer, una stringa o una sequenza disordinata di caratteri nella quale cercare un ordine. Un sistema di zipping, infatti, altro non fa che diminuire il disordine, o entropia. La chiave di lettura sta nel misurare l'entropia relativa: attraverso il confronto di file zippati il computer è in grado di individuare le peculiarità degli stili letterari e di una lingua. Più le lingue o gli stili letterari a confronto sono simili, maggiore sarà l'efficacia di zipping e minore l'entropia relativa che si attribuisce a un testo rispetto a un altro di riferimento.

I ricercatori hanno fatto leggere al computer ben 90 testi di 11 famosi autori italiani, tra cui Dante, Machiavelli, Manzoni e Pirandello. Successivamente gli hanno sottoposto testi ignoti e il computer ne ha indovinato l'autore con un successo del 93 per cento. Ma c'è di più: dopo aver analizzato dieci testi in altrettante lingue, al computer sono bastati 20 caratteri per stabilire in quale lingua un qualunque testo è scritto. Una macchina esperta di letteratura italiana, poliglotta e anche filologa. Dopo l'analisi di oltre 50 versioni di quello che è ritenuto il documento più tradotto al mondo, la Dichiarazione Universale dei Diritti Umani, il computer ha ricostruito le relazioni filologiche tra le lingue, riconoscendo persino la singolarità della lingua maltese e di quella basca.

"Il metodo da noi proposto", spiega Dario Benedetto, ricercatore del team, "è estremamente generale e versatile e può, in linea di principio, trattare qualunque sequenza di caratteri, da un testo letterario alle sequenze di Dna. Le nostre ricerche si sono finora concentrate su problemi in ambito linguistico e sull'analisi di sequenze temporali; esperimenti preliminari sono stati condotti su sequenze di Dna che codificano proteine". Le applicazioni sono molte, dalla genetica all'economia, alla previsione di catastrofi naturali. "Mentre per un testo si è interessati al riconoscimento di proprietà come l'argomento trattato o l'autore", continua Benedetto, "per una sequenza temporale si vorrebbe riuscire a individuare specifiche sottosequenze che segnalino il sopraggiungere di eventi estremi, come grossi tracolli in borsa o violenti terremoti".

Ma sarà in grado il computer di smascherare un ghost writer, l'anonima penna che si nasconde dietro una firma fittizia? "Sì, ma con le dovute cautele", afferma Emanuele Caglioti, ricercatore dell'équipe romana, "siamo già impegnati in alcune sperimentazioni in questa direzione, ma ovviamente non si può pretendere di ottenere risultati incontrovertibili. Si potrebbero avere chiare indicazioni di paternità di un testo, ma nulla vieta che un autore possa scrivere esattamente alla maniera di un altro e ingannare il metodo. Uno degli esperimenti che abbiamo in programma riguarda l'analisi di testi scritti parafrasando autori noti".

"Ogni autore ha certamente dei caratteri distintivi nell'uso delle parole, nella costruzione dei periodi o nell'uso della punteggiatura, tutti individuabili da un letterato esperto", sottolinea Vittorio Loreto, altro componente del team di ricerca, "potrebbero però esistere elementi distintivi di cui è difficile avere consapevolezza. La macchina, in questo caso, potrebbe essere in grado di leggere, con un occhio perfettamente obiettivo, i dettagli apparentemente più insignificanti ma forse cruciali, elementi, per così dire, invisibili. Ma l'apprezzamento del bello, comunque, e per fortuna, è sicuramente fuori dalle nostre potenzialità di ricerca".