

Il legame big data - intelligenza artificiale nel contrasto alla corruzione

Author : Anna Cardetta

Date : 15 Gennaio 2019



Si terrà a breve a Parigi il Global Anti-Corruption & Integrity Forum 2019 dell'OCSE^[1], consueto appuntamento annuale dedicato a tematiche in ambito internazionale relative a corruzione, governance, concorrenza e cooperazione, al quale prenderanno parte leader di governo, imprenditori, esponenti della società civile ed esperti con il comune obiettivo di promuovere e rendere sempre più efficaci iniziative a sostegno del contrasto al malaffare e alla corruzione. L'attenzione del Forum 2019 si focalizzerà sui rischi e sulle opportunità che caratterizzano le nuove tecnologie che stanno trasformando il modo di lavorare di governi, imprese e società, riferendosi in particolar modo a blockchain, big data analytics e intelligenza artificiale.

Senza pretesa alcuna di argomentare su ciò che sarà oggetto di autorevoli interventi nel corso dell'importante evento e tralasciando in questa analisi le considerazioni sulla tecnologia blockchain come strumento anticorruzione, proviamo a trarre qualche considerazione sull'interessante legame big data - intelligenza artificiale e su come questo possa in effetti rappresentare un reale ed efficace metodo di contrasto alle dinamiche corruttive, soprattutto in un'ottica preventiva di intercettazione dei reati.

Gli open data diventano big

Non vi è dubbio che la condivisione delle informazioni tra Stati attraverso gli open data^[2] contribuisca in maniera determinante ad identificare lacune critiche in ambito anticorruzione e ad incentivare la collaborazione a livello globale. L'obiettivo generale è quello di consentire un uso ampio ed esteso dei dati aperti delle amministrazioni pubbliche da parte di chiunque e ovunque, senza limitazioni giuridiche o tecniche (ad esempio, i diritti d'autore, formati proprietari) e con la dovuta osservanza dei requisiti di sicurezza e privacy.

I dati aperti agevolano la prevenzione e il contrasto della corruzione, responsabilizzando gli Stati e facendo chiarezza sulle attività dei governi, sulle decisioni e sulle spese sostenute, consentendo ai cittadini di monitorare il flusso e l'uso del denaro pubblico all'interno dei Paesi di appartenenza, permettendo il confronto dei dati con quello di altri Paesi e il monitoraggio del

flusso di fondi pubblici a livello transfrontaliero.

I dati aperti sono in grado di fornire alle aziende informazioni in tempo reale sul contesto economico di un determinato Paese, mercato o settore, consentendo alle stesse consapevoli e ponderate decisioni di investimento e contribuiscono alla costruzione di sistemi organizzativi e contesti lavorativi più aperti, trasparenti e meno corrotti, riducendo il rischio di conflitti di interesse e favoritismi.

La produzione da parte delle PA (e l'analisi da parte dei governi) di grandi quantità di dati amministrativi e dataset finalizzati a descrivere e tracciare processi e attività ritenuti a rischio di corruzione, è divenuta nel tempo sempre più orientata ad un approccio di tipo 'big data'[\[3\]](#) basato sulla elaborazione, combinazione, ed analisi di grandi quantità di dati provenienti da fonti eterogenee.

L'applicazione di adeguate tecniche di analisi su tali quantità di dati può fornire una serie di opportunità per la definizione di più solide politiche anticorruzione e per la valutazione di prestazioni e servizi in termini di spesa pubblica e di comportamenti organizzativi. Ad esempio, il collegamento dei dati in materia di appalti pubblici con altri insiemi di dati amministrativi riguardanti le aziende private operanti sul territorio possono fornire indicatori quantitativi e qualitativi per descrivere la trasparenza operativa, la capacità amministrativa e la qualità dell'operato di PA e imprese a vari livelli di analisi.

L'importanza dell'Intelligenza Artificiale per l'analisi dei big data

Intelligenza artificiale[\[4\]](#) e big data sembrano rappresentare l'ottimale impiego della tecnologia informatica nella lotta alla corruzione: tramite l'uso delle reti neurali[\[5\]](#) e degli strumenti di apprendimento automatico[\[6\]](#) è possibile effettuare analisi sui dati a un tale livello di granularità da poter individuare relazioni, collegamenti e anomalie all'interno di grandi dataset.

Nel gennaio 2018 ricercatori della Higher School of Economics (HSE) e dell'università spagnola di Valladolid, partendo dal presupposto che i fenomeni corruttivi debbano essere necessariamente rilevati quanto prima in modo da adottare idonee misure atte a contrastarli, hanno sviluppato un sistema di allerta basato sull'utilizzo di reti neurali, in particolare mappe auto-organizzanti[\[7\]](#), in grado di intercettare la corruzione in ambito pubblico basandosi sull'analisi di fattori economici e politici.

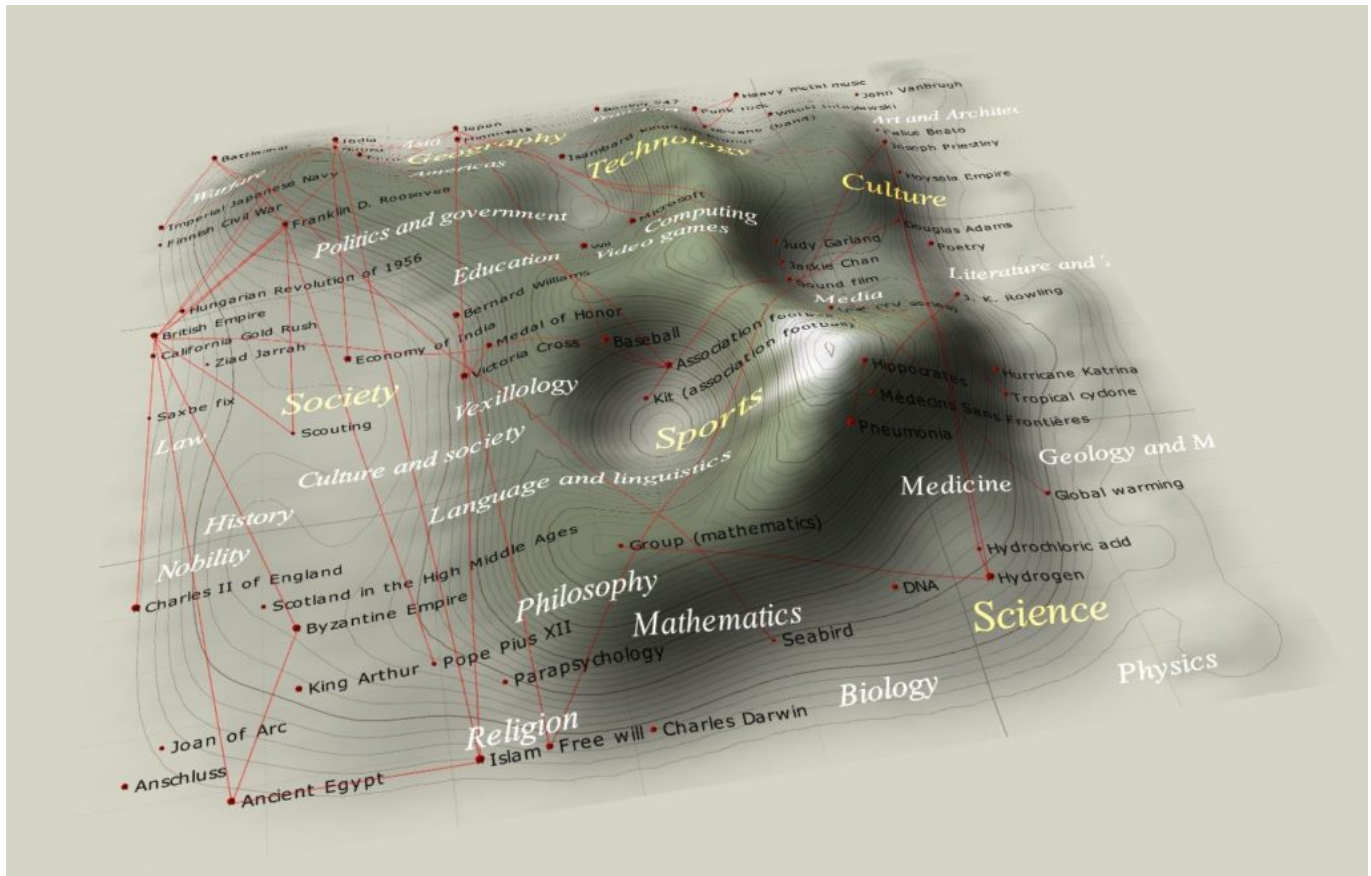


Figura 1 - Rappresentazione cartografica di una mappa auto-organizzante (U-Matrix) basata su dati di articolo in evidenza su Wikipedia (frequenza delle parole). La distanza è inversamente proporzionale alla somiglianza. Le 'montagne' sono spigoli tra i cluster. Le linee rosse sono collegamenti tra articoli (Fonte: Wikipedia)

Nello specifico, il sistema di intelligenza artificiale implementato è in grado di individuare le province spagnole che corrono maggiori rischi di corruzione e di identificare le variabili economiche e politiche che inducono la corruzione pubblica (ad es. la tassazione immobiliare e l'aumento dei prezzi degli immobili, il permanere al potere dello stesso partito politico per lunghi periodi di tempo e così via).

A seconda delle caratteristiche di ciascuna regione analizzata, secondo i ricercatori è possibile stimare la probabilità di casi di corruzione emergenti in un periodo di tre anni. Mentre in alcune circostanze i casi di corruzione possono essere previsti ben prima che si verifichino e consentire quindi l'attuazione di misure preventive, in altri casi il periodo di previsione è molto più breve, rendendo necessarie urgenti misure correttive. Il modello può essere applicato anche ad altre regioni e Paesi e potrebbe essere migliorato se si prendessero in considerazione fattori specifici per singolo Paese o regione. Seguendo la guida dei ricercatori spagnoli, dunque, i governi potrebbero utilizzare i sistemi di intelligenza artificiale per identificare le vulnerabilità e indirizzare le azioni e i controlli in particolari aree a rischio.

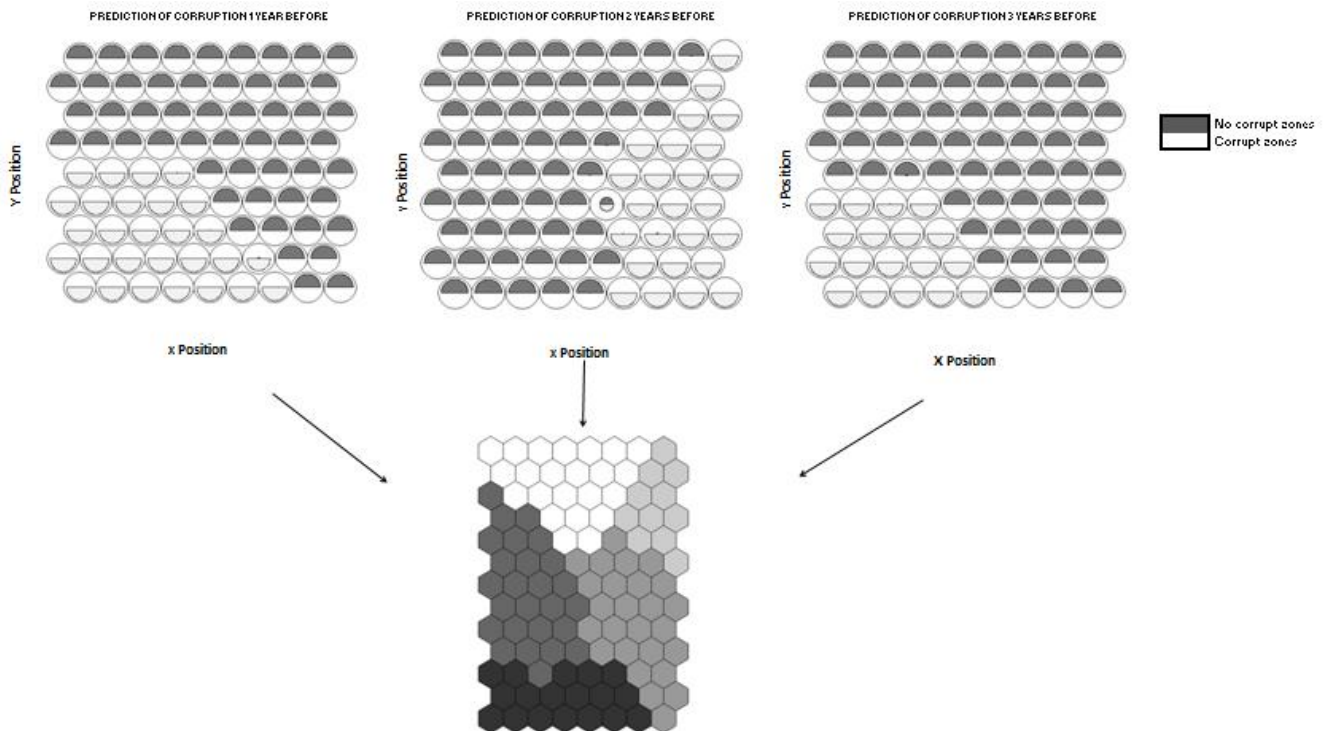


Figura 2 – ‘Predicting public corruption with neural networks: an analysis of Spanish provinces’, rappresentazione visuale del modello - (Félix J. López-Iturriaga, University of Valladolid, Higher School of Economics - Iván Pastor Sanz, University of Valladolid)

Il punto di forza dei sistemi di modellazione basati su reti neurali è dato dall’essere strumenti potenti e flessibili che consentono di superare il limite di dover fare ipotesi restrittive sul processo di generazione dei dati o sulle leggi statistiche riguardanti le variabili rilevanti. Questi sistemi sono caratterizzati dalla capacità di elaborare grandi quantità di dati a diversi livelli di dettaglio; questo consente di poter accelerare le attività di analisi dei dati di routine e di poter disporre di una maggiore disponibilità di risorse umane per attività di approfondimento e controllo su operazioni sospette o da monitorare nel tempo.

L’intelligenza artificiale rappresenta un insieme di tecnologie capaci di *apprendere* piuttosto che affidarsi alle istruzioni impartite dagli sviluppatori e di fare previsioni al di là della normale capacità umana. A livello aziendale, ad esempio un sistema di intelligenza artificiale potrebbe *capire* il contenuto di normative e regolamenti e imparare a riconoscere i modelli aziendali anticorruzione ‘conformi’ a tali disposizioni; eventuali revisioni successive delle leggi e dei regolamenti (o altre modifiche rilevanti) potrebbero essere direttamente incorporate dal sistema di intelligenza artificiale senza alcun intervento umano.

Conclusioni

Il ricorso all’intelligenza artificiale consente di ridurre sensibilmente i compiti di routine, ma le risorse umane sono importanti per indirizzare il sistema di intelligenza artificiale nella giusta

direzione e per incanalare le informazioni 'supervisionate' quando il sistema viene sviluppato per la prima volta.

Il risultato della ricerca spagnola non rappresenta il primo caso di impiego della tecnologia informatica in ambito anticorruzione, dato che già da tempo governi, organizzazioni internazionali e società civile sono impegnate nella diffusione e nel perfezionamento di strumenti e metodologie di raccolta, tracciamento e analisi di informazioni e dati per il contrasto dei fenomeni corruttivi: ciò che distingue l'innovazione della ricerca spagnola è l'utilizzo dell'intelligenza artificiale che consente al modello di poter identificare la corruzione prima che emerga, rappresentando un valido supporto alle autorità governative nell'individuazione delle misure preventive da adottare o rimodulare.

L'auspicio è quello di poter addestrare algoritmi su set di dati sempre più complessi ed eterogenei alla ricerca di alert e indicatori di corruzione sempre più difficili da individuare e che Università e centri di ricerca continuino ad essere coinvolte assiduamente nella progettazione e nella implementazione di soluzioni innovative che possano agevolmente essere utilizzate nella quotidiana lotta al contrasto dei fenomeni corruttivi.

Note

[1] <http://www.oecd.org/corruption/integrity-forum/>

[2] 'L'**open data** si richiama alla più ampia disciplina dell'open government, cioè una dottrina in base alla quale la pubblica amministrazione dovrebbe essere aperta ai cittadini, tanto in termini di trasparenza quanto di partecipazione diretta al processo decisionale, anche attraverso il ricorso alle nuove tecnologie dell'informazione e della comunicazione, e ha alla base un'etica simile ad altri movimenti e comunità di sviluppo "open", come l'open source, l'open access e l'open content' - Fonte: Wikipedia

[3] 'Il termine **big data** ("grandi [masse di] dati" in inglese) descrive l'insieme delle tecnologie e delle metodologie di analisi di dati massivi, ovvero la capacità di estrapolare, analizzare e mettere in relazione un'enorme mole di dati eterogenei, strutturati e non strutturati, per scoprire i legami tra fenomeni diversi e prevedere quelli futuri' - Fonte: Wikipedia

[4] 'L'intelligenza artificiale è una disciplina recente che negli anni ha fornito un importante contributo al progresso dell'intera informatica. Essa è stata inoltre influenzata da numerose discipline fra le quali la filosofia, la matematica, la psicologia, la cibernetica, le scienze cognitive. L'intelligenza artificiale studia i fondamenti teorici, le metodologie e le tecniche che consentono di progettare sistemi hardware e sistemi di programmi software atti a fornire all'elaboratore elettronico prestazioni che, a un osservatore comune, sembrerebbero essere di pertinenza esclusiva dell'intelligenza umana. Suo scopo non è quello di replicare tale intelligenza, obiettivo che per taluni è addirittura non ammissibile, bensì di riprodurre o emularne alcune funzioni. Non vi è alcun motivo che impedisca a priori che talune (ma non tutte) prestazioni dell'intelligenza umana – per esempio la capacità di risolvere problemi mediante processi inferenziali – possano essere fornite anche da una macchina. Nel caso dell'emulazione, le prestazioni intelligenti sono ottenute utilizzando meccanismi propri della macchina, in modo da fornire prestazioni

qualitativamente equivalenti e quantitativamente superiori a quelle umane.' - Fonte: Francesco Amigoni, Viola Schiaffonati, Marco Somalvico Treccani, Enciclopedia delle Scienza e della Tecnica, 2008

[5]'Una rete neurale è un insieme di neuroni biologici tra loro interconnessi. Nell'uso moderno si intende però di solito con rete neurale una rete di neuroni artificiali, che cerca di simulare il funzionamento dei neuroni all'interno di un sistema informatico. Può essere composta sia da programmi che da hardware dedicato. Spesso viene utilizzata in congiunzione alla logica fuzzy e funge da base per molti sistemi di intelligenza artificiale.' - Fonte: Wikibooks

[6]'L'apprendimento automatico (noto anche come machine learning) rappresenta un insieme di metodi sviluppati a partire dagli ultimi decenni del XX secolo in varie comunità scientifiche con diversi nomi come: statistica computazionale, riconoscimento di pattern, reti neurali artificiali, filtraggio adattivo, teoria dei sistemi dinamici, elaborazione delle immagini, data mining, algoritmi adattivi, ecc; che utilizza metodi statistici per migliorare progressivamente la performance di un algoritmo nell'identificare pattern nei dati.' - Fonte: Wikipedia

[7]'Il concetto di mappa auto-organizzante (SOM, Self organizing map) si basa sul principio dell'apprendimento competitivo, in cui un elemento della rete (un nodo o un neurone nel modello cerebrale) si specializza nel riconoscere uno stimolo (vince la competizione con gli elementi vicini). L'evoluzione temporale dell'apprendimento definisce una configurazione stabile (topologia di uscita) degli elementi coordinati dal nodo vincitore, secondo pesi variabili (tipicamente con la distanza). Si stabilisce in definitiva un isomorfismo tra spazi di ingresso e di uscita, ossia tra stimoli (input) e risposta standardizzata agli stessi.' Fonte: Fabio Catino - Treccani, Enciclopedia delle Scienza e della Tecnica, 2008

Articolo a cura di **Anna Cardetta**