

# Quando il bibliotecario migliore è un'IA

Come annunciato nell'editoriale del n. 1/2020 il rafforzamento dell'approccio internazionale della rivista è uno degli obiettivi che intendiamo perseguire, attraverso un contatto costante con il mondo della professione e attraverso un consolidamento del rapporto con il comitato scientifico internazionale. Di questo fa parte ora, con nostro grande piacere, David Weinberger, al quale abbiamo chiesto di proporre ai lettori di *AIB studi* una riflessione su un tema di forte attualità bibliotecaria. Il tema prescelto è quello del ruolo dell'IA (intelligenza artificiale) e del *machine learning* (in italiano apprendimento automatico) nei servizi di 'raccomandazione bibliografica' esterni e interni alle biblioteche.

Weinberger offre il suo punto di vista a partire dall'esperienza acquisita nel ruolo di co-direttore dell'Harvard Library Innovation Lab nonché a seguito dell'attività svolta per il Google's People + AI Research (PAIR), un team multidisciplinare di Google che esplora il lato umano dell'IA, a partire dall'idea che per raggiungere il suo potenziale positivo, il *machine learning* debba essere partecipativo, coinvolgendo le comunità di interesse. È alla comunità dei bibliotecari che Weinberger si rivolge richiamandoli ad avere un ruolo proattivo nell'utilizzo dell'IA al servizio delle proprie comunità di utenti.

L'IA – come tutte le tecnologie – è neutra rispetto ai valori e agli scopi di utilizzo: per questo Weinberger ritiene che le biblioteche debbano fare da contraltare rispetto a un suo utilizzo puramente commerciale al fine di infondere nei meccanismi del *machine learning* quei valori «democratici e umanitari» di cui le biblioteche sono veicolo e che in questo momento storico richiedono una rinnovata attenzione. Per fare questo i bibliotecari non devono arrendersi di fronte alla complessità e devono puntare a rendere interoperabili gli strumenti che hanno costruito negli anni e valorizzare le competenze acquisite. In questo senso è necessario sempre più ragionare in un'ottica sistemica superando la frammentazione che è forse il vero punto debole delle biblioteche al cospetto dei competitor commerciali.

Una sfida affascinante che speriamo anche i bibliotecari italiani – in un'epoca come quella che stiamo vivendo – vogliano raccogliere per il futuro delle nostre biblioteche e delle nostre comunità di riferimento.

*Chiara Faggiolani, Anna Galluzzi*

\*\*\*\*

Fino a ora siamo stati in grado di liquidare con una scrollata di spalle la sfida che l'IA (intelligenza artificiale) pone alle biblioteche, insistendo che gli uomini siano in grado di prendere, sulle persone e sulla cultura, decisioni migliori e molto più ricche di sfumature di quanto possa fare una qualunque macchina senza cuore.



Abbiamo poi limitato quella dichiarazione, dato che l'IA si sta dimostrando utile laddove le statistiche possano offrire un significativo supporto alle decisioni: per generare liste di documenti candidati alla rimozione da scaffale, al limite per prevedere quali potrebbero registrare un'impennata di popolarità, e certamente per massimizzare l'efficienza nella gestione degli spazi per l'archiviazione.

Tuttavia, quando si tratta delle decisioni che i bibliotecari prendono sulla base dell'esperienza e della saggezza che hanno costruito nel tempo, del loro impegno per lo sviluppo delle loro comunità, alcuni continuano a insistere che ci sia un cerchio magico che protegge i bibliotecari – una zona ove ci sia un 'Divieto per le macchine!'.

Comprensibile, ma rischioso. Sembra che ogni giorno saluti la notizia di nuovi successi delle macchine nello svolgere compiti che consideravamo totale appannaggio dell'intelletto umano, della creatività e dell'emozione. Non sappiamo ancora quanto sarà ampio il cerchio magico alla fine del processo, ma è verosimile che continuerà a restringersi.

Alcune aree sembrano più al riparo dalla diffusione dell'IA. Per esempio, se il ruolo dello spazio fisico si evolve, le biblioteche resteranno probabilmente luoghi insostituibili, sicuri, utili e tranquilli per la comunità.

Allo stesso modo è improbabile che le biblioteche soccombano di fronte all'IA in quanto fonti di libero accesso ai materiali culturali. Le sfide possono giungere da altre direzioni – per esempio, la legge sul copyright potrebbe un giorno fare i conti con il modo in cui internet ha liberato la cultura – ma ciò appare scollegato dall'IA.

Poi c'è il semplice fatto che per la gran parte delle comunità l'edificio della biblioteca costituisce segno e simbolo del loro attaccamento alla cultura, all'apprendimento e alla giustizia.

Ma l'IA sta già lanciando la sfida a una delle principali virtù delle biblioteche: una biblioteca, considerata come sistema, accumula grandi conoscenze sia sugli oggetti dei quali si prende cura che sulle comunità che serve; un recente articolo in *Code4Lib* di Greg Sohanchyk e Dan Briem descrive in dettaglio i metodi per espandere quella conoscenza relativamente alla propria comunità. Ecco perché gli utenti si rivolgono alle biblioteche e ai bibliotecari per avere consigli sulla prossima cosa che dovrebbero leggere, vedere o ascoltare.

Immagino che siamo molto distanti dal desiderare di chiacchierare con una IA come fosse un bibliotecario umano, godendoci lo scambio di battute e sentendoci come se stessimo facendo conoscenza. Ma quando si tratta della parte transazionale della conversazione – l'effettivo consiglio – l'IA ha dei vantaggi significativi sugli umani. Tanto per cominciare, l'IA è in grado di immagazzinare molte più informazioni sui libri e sui singoli utenti di quante possano entrare nella testa di un qualunque essere umano.

Ad esempio, Amazon conosce i dettagli di milioni di libri. Sa su quali modelli si basano le scelte che gli utenti compiono per ciascuno di essi. È al corrente di come si intersecano le reti delle connessioni sociali reali e potenziali degli utenti. Vede il comportamento di acquisto, i clic e le recensioni dei suoi utilizzatori, e non solo per quanto riguarda i libri. Sa di quali titoli gli utenti hanno scaricato un estratto sui loro Kindle e quali scaricamenti hanno condotto a una vendita. Potrebbe facilmente distinguere quali libri gli utenti, sui loro Kindle, hanno iniziato ma non finito. In termini di dati puramente quantitativi, nessun bibliotecario ha mai memorizzato nemmeno una frazione di ciò che Amazon sa su libri e lettori.

Ma vi sento obiettare che Amazon è un'impresa spietatamente capitalista che è interessata solo alla vendita di un certo prodotto ai propri utenti. Esattamente così.

Questo offre alle biblioteche la loro occasione. E se non la colgono, l'IA nelle mani di entità commerciali costituisce la maggior minaccia per le biblioteche dopo la fisica del fuoco, poiché l'utilità dell'IA fornita da compagnie in cerca di profitto continuerà a crescere, lasciando che le biblioteche mangino la polvere.

Le biblioteche hanno due cose dalla loro.

Primo, la gente in genere crede che i bibliotecari sappiano di più sui libri di quanto possa sapere una qualunque macchina, ma ho paura che non durerà. Non solo un sistema di *machine learning* è in grado di registrare informazioni su milioni di libri e altre risorse – inclusa l'intera coda lunga delle opere raramente lette o consigliate – ma le società di lucro hanno forti motivazioni ad addestrare i sistemi di *machine learning* (o apprendimento automatico) affinché combinino gli interessi delle persone in modo ancora più accurato.

La seconda ragione è più forte: le biblioteche sono dalla parte dei loro utenti e delle comunità. Ad Amazon non importa quale libro compriamo fin tanto che ne acquistiamo uno, quindi i suoi algoritmi danno 'consigli orientati al minimo sforzo'. Se Amazon vede che abbiamo comprato libri sui politici di sinistra o di destra, sul cambiamento o sul negazionismo climatico, su una religione e non altre, ci offrirà tanti più libri simili a quelli già acquistati quanti ne può trovare.

Questo accade perché il *machine learning* individua schemi, mentre ai bibliotecari piace romperli, gli schemi. Se avete appena completato una serie di novelle romantiche ambientate nell'Inghilterra del XIX secolo, la vostra bibliotecaria vi informerà che esiste una serie simile, o altre opere dello stesso autore. Ma potrebbe anche menzionare una novella di Jane Austen o forse un saggio biografico su una donna vissuta in quell'epoca. Amazon vuole solamente che compriate un libro, ma i bibliotecari usano il loro discernimento per spingervi ad ampliare i vostri interessi. Essere dalla parte degli utenti spesso significa espandere la loro curiosità, allargare un po' la loro visione.

Questo significa anche che le biblioteche stanno dalla parte delle comunità aperte e tolleranti. Amazon, d'altro canto, sa che è più facile indurci ad acquistare un libro che conferma le nostre credenze piuttosto che uno che le mette alla prova.

Quindi, come possono le biblioteche aiutare a smussare un ambiente culturale nel quale il prossimo oggetto culturale con il quale entrerete in relazione è probabile che venga accuratamente e proficuamente diretto su di voi da un'IA al servizio di un'azienda che, a differenza delle biblioteche, non si cura di ampliare gli orizzonti delle persone o di creare una comunità meglio informata, dalla mente più aperta, più simpatica, tollerante e compassionevole?

Potremmo cambiare l'infrastruttura della nostra biblioteca.

Può suonare esagerato e costoso, e lo sarebbe se realizzato al livello massimo. Ma potrebbe anche essere fatto in modo incrementale, e porterebbe di certo benefici per le biblioteche locali.

Lasciate che vi offra un quadro di massima, e in seguito parleremo dei piccoli passi.

La soluzione più ovvia ai problemi posti dalle società di profitto è costruire un'alternativa non commerciale. Per il momento, chiamiamola il 'grafo della biblioteca aperta', *open library graph*; un grafo è una rappresentazione densamente interconnessa di quantità enormi di dati disparati. Immaginate un nodo per l'Inferno di Dante che abbia connessioni a tutti i libri che fanno riferimento all'Inferno, ai libri accademici su Inferno e Paradiso, a informazioni sui guelfi bianchi, sui guelfi neri, sui ghibellini, sul *Giardino delle delizie* di Bosch e da lì al moderno surrealismo, alla Bibbia cristiana, alla Bibbia ebraica, al Corano, al fato di quelli, come Socrate, nati prima di Cristo ma degni del Paradiso, poi ancora alle opere della filosofia occiden-

tale, della filosofia orientale – una concatenazione interconnessa di lavori culturali e citazioni, che si espande continuamente e gradualmente si arricchisce. Costituirebbe una risorsa globale di valore incalcolabile, e attingerebbe a dati già disponibili globalmente: Europeanana, la Digital Public Library of America, i sistemi bibliotecari nazionali di molte nazioni, i dati bibliografici delle università ecc.

Le biblioteche dovrebbero costruire e possedere questa risorsa aperta.

Ed essa dovrebbe apprendere da ciò che le biblioteche sanno sulle loro comunità, mentre naturalmente preserverebbe l'anonimato dei singoli utenti.

Ecco un esempio. Quando ero condirettore dell'Harvard Library Innovation Lab, avevamo elaborato uno *'stackscore'* per ogni opera del catalogo universitario, applicando una formula pesata che considerava dati aggregati ma resi anonimi, come la frequenza con la quale l'opera veniva presa in prestito, quella con la quale veniva messa da parte per i corsi e altre simili misurazioni. Il risultato era un numero tra 1 e 100 per ciascun libro, che serviva come misura grezza ma utile del giudizio della comunità di Harvard su quella risorsa (questi calcoli hanno bisogno di alcuni passi aggiuntivi per assicurare l'anonimato). Poi scrivemmo un browser per il catalogo che ordinasse i risultati sulla base del loro *stackscore*, sfruttando l'uso che la comunità faceva dei materiali per individuare le opere più utili.

Una delle cose buone di *stackscore* è che ogni biblioteca potrebbe elaborare una propria formula per calcolare i valori della comunità di riferimento. Probabilmente alcune darebbero maggior peso alla frequenza con cui un libro viene richiesto, mentre per altre potrebbe contare di più il numero di proroghe richieste, o i consigli del bibliotecario di reference. Così, le misure del valore potrebbero essere locali mentre lo *stackscore* – un numero tra 1 e 100 – permetterebbe di fare confronti tra diverse biblioteche.

In effetti, sarebbe plausibile usare le *learning machine* per calcolare *stackscore* più accurati e funzionali. L'analisi potrebbe scendere maggiormente nei dettagli e, ciò che più conta, potrebbe essere possibile predire degli *stackscore* per oggetti finiti tanto in fondo alla coda lunga da non essere stati letti per decenni, valorizzando così la collezione della biblioteca.

Lo *stackscore* è solo un'idea. Ma questo genere di idee, per alcune ragioni precise, ha un valore speciale se confrontato con le classifiche commerciali.

Primo, esso risponde alle preferenze personali con le preferenze di comunità, valorizzando le comunità locali proprio in riferimento alla loro specificità.

Secondo, qualsiasi sistema di raccomandazione basato su dati relativi alla storia dell'uso dei materiali verosimilmente crea un *feedback* circolare: se un libro viene consigliato a causa del suo alto utilizzo, ogni suggerimento finirà per aumentare quello stesso uso. Tuttavia, se esistesse un'infrastruttura condivisa per i dati bibliotecari, una comunità potrebbe sfruttare abitualmente gli *stackscore* di altre comunità per indirizzare le persone verso opere che altrimenti potrebbero sfuggire loro. Per esempio: 'Questi sono i dieci libri sul tema dell'evoluzione che la nostra comunità ha maggiormente utilizzato e apprezzato. Ma questi sono i primi dieci nell'università cittadina, al dipartimento di biologia e al dipartimento di teologia, e questi sono i preferiti di una comunità culturalmente distinta dalla nostra'.

Calcolare e condividere gli *stackscore* è relativamente facile. Ma tutti gli aspetti delle comunità che imparano l'una dall'altra, e in fin dei conti forse contribuiscono a qualcosa come un 'grafo della biblioteca aperta', verrebbero di molto facilitati da software di gestione delle biblioteche che facilitino la condivisione dell'informazione. Il metodo per farlo dal basso sarebbe far pressione sui fornitori dei sistemi integrati per le biblioteche affinché rendano quel tipo di informazione disponibile

in formati e protocolli standard. Il metodo per realizzarlo dall'alto sarebbe di mettere finalmente a disposizione nel *cloud* i servizi ILS, gratuitamente o a prezzo molto basso, e a codice aperto, così da rendere disponibile a livello globale l'informazione aggregata sulle collezioni e sulle abitudini – allo stesso tempo preservando con cura la riservatezza degli utenti. Questa costituirebbe una fonte di dati fantastica per le biblioteche o altri soggetti che desiderassero usare applicazioni di *machine learning* a supporto degli utenti e delle loro comunità locali.

Tutti questi suggerimenti specifici sono semplicemente esempi. Potrebbero non essere fattibili o nemmeno desiderabili, ma in un modo o nell'altro le biblioteche devono aprirsi alle possibilità offerte dal *machine learning* per supportare i loro utenti e le loro comunità. Abbandonare questo campo alle società commerciali pone a rischio la cultura stessa. Dipende dalle biblioteche grandi e piccole umanizzare il *machine learning* per assicurare che si metta al servizio di quei valori profondamente democratici e umanitari che sono propri anche della più umile delle biblioteche pubbliche.

*David Weinberger*