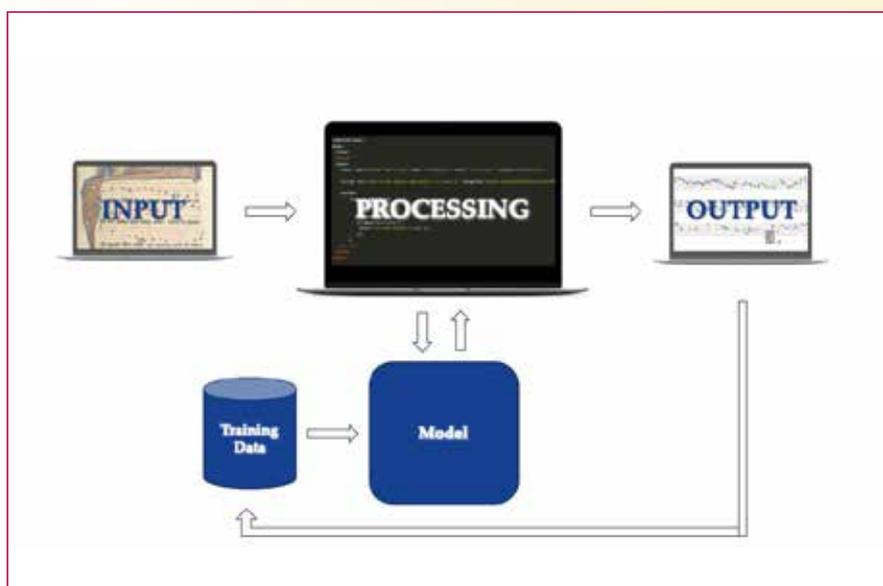


HANDWRITTEN TEXT AND MUSIC RECOGNITION AS A RESOURCE FOR CULTURAL HERITAGE

di Roberto Ribuoli



Le tecnologie di riconoscimento del testo manoscritto, Handwritten Text Recognition, così come quelle di riconoscimento della notazione musicale, Optical Music Recognition, nell'ultimo quinquennio hanno accelerato la propria evoluzione, trascinate dall'affermazione dell'Intelligenza Artificiale e dal miglioramento delle proprietà di calcolo e archiviazione dei computer.

La loro applicabilità su scala rappresenta una risorsa cruciale per il patrimonio culturale, in quanto capace di offrire una connessione solida nel punto di discontinuità tra bene materiale ed immateriale. Presentiamo in quest'articolo il loro funzionamento e le prospettive di utilizzo, dialogando con Günter Mülberger e Andy Stauder di Transkribus, software leader nell'HTR per documenti storici, e Jorge Calvo-Zaragoza, una delle voci più autorevoli nell'OMR per la ricerca.

HTR e OMR condividono, pur nella diversità degli oggetti di analisi, un'architettura analoga, articolata in input, processing e output. Al sistema viene presentata un'immagine di manoscritto, l'input, che, processata, viene restituita in un formato codificato (output) che conserva le informazioni significative dell'originale.

Entrambe le tecnologie si basano, all'attuale stato dell'arte, sul Deep Learning e dipendono da dati annotati sui quali il sistema viene allenato. Nel nostro caso, tale base dati è costituita da immagini di pagine corredate della loro forma codificata: il processo di annotazione deve essere realizzato, o supervisionato perlomeno, da un esperto. In questo modo il sistema "impara" a leggere documenti con scritture o notazioni analoghe, migliorando la sua efficacia all'aumento della quantità dei dati, che sono generati ad ogni nuovo utilizzo. Come in altri settori dell'AI, è importante ricordare che i processi definiti apprendimento e comprensione da parte del sistema consistono più propriamente in un'affinamento progressivo di una curva probabilistica: gli elementi in input vengono ricondotti alla cate-

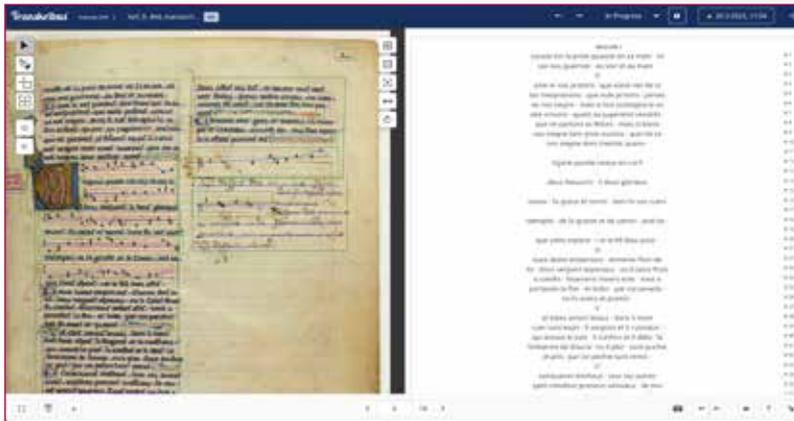
ria più plausibile disponibile nel modello allenato, e il perfezionamento del modello corrisponde ad una percentuale maggiore di assegnazioni corrette.

HTR e OMR hanno suscitato l'interesse della comunità scientifica per le loro applicazioni in ricerche storico-linguistiche e musicologiche di diverse dimensioni, come supporto all'indicizzazione e all'analisi degli oggetti di studio. Nel 2023 grazie all'HTR è stato individuato un dramma attribuibile a Lope de Vega¹ in un fondo della Biblioteca Nazionale di Spagna, è stata provata definitivamente la falsità dei cosiddetti Diari di Hitler², e grazie alla disponibilità di una tecnologia OMR avanzata è stato avviato l'ambizioso progetto REPERTORIUM, che si propone di digitalizzare e rendere accessibile l'intero corpus del canto gregoriano³. Questi contributi alla ricerca costituiscono però solo un lato delle potenzialità di HTR e OMR che, per quanto nate in un diverso contesto, possono essere considerate oggi a pieno titolo risorse per la valorizzazione digitale del patrimonio culturale. La digitalizzazione resa possibile da HTR e OMR è una digitalizzazione di seconda generazione, in quanto non si limita al confezionamento di una rappresentazione digitale dei documenti, ma corredata tale rappresentazione di un'interpretazione del suo contenuto: all'immagine di un manoscritto scansionato sono ancorate, in un file abbinato, informazioni relative al testo contenuto, al tipo di scrittura, alla plausibile datazione. Questo consente di tutelare pienamente il valore culturale immateria-

le del documento, il significato che trasmette, e non solo, attraverso la sua scansione digitale, il documento in quanto bene materiale. Se, come suggerisce Lorenzo Casini, *l'immaterialità è connaturata ai beni culturali, anche perché immateriale e intangibile è, al fondo, il motivo più alto che spinge a proteggerli e a favorirne la conoscenza*⁴, una tutela dei beni che si arresti a una digitalizzazione muta, alla mera copia del documento, non ne tutela allora pienamente il valore. HTR e OMR offrono una risorsa per risolvere la discontinuità tra materiale e immateriale che si crea nella digitalizzazione di prima generazione e risultano particolarmente utili nei casi in cui il valore culturale dei beni conservati sia *diffuso*. Cosa si intende qui per diffuso? Un manoscritto che conserva l'unica testimonianza di un testo classico concentra in sé un valore importante, irrelato rispetto ad altri oggetti, a bassa diffusione; un archivio di libri contabili seicenteschi di un comune italiano ha invece un valore diffuso, in quanto ogni libro preso singolarmente ha una bassa importanza, ma il loro insieme conserva informazioni rilevanti per la ricostruzione di quel periodo storico. Un autografo di Bach ha un valore culturale poco o per nulla diffuso; il corpus di composizioni polifoniche liturgiche di un archivio diocesano poco indagato ha un valore diffuso, perché è da un'analisi d'insieme che si possono evincere informazioni significative sulla pratica musicale testimoniata dai componimenti, singolarmente pur di limitato valore.

HTR e OMR permettono dunque di agevolare la tutela del patrimonio culturale in situazioni di diffusione complesse: in molti casi infatti estrarre manualmente informazioni dai singoli documenti – trascriverli, annotarli – ha un costo alto e difficilmente finanziabile. L'effettiva messa in atto delle potenzialità di queste tecnologie dipende però da come verranno risolte questioni ancora aperte: la definizione di standard e la condivisione dei dati, il coinvolgimento della comunità accademica, la cooperazione tra istituzioni private e università. Su tali questioni ho dialogato, in momenti diversi, con Günter Mülberger e Andy Stauder per l'HTR, e con Jorge Calvo-Zaragoza per l'OMR. Stauder e Mülberger sono rispettivamente managing director e chair nel board di ReadCOOP, società cooperativa che sviluppa il software HTR Transkribus, Calvo-Zaragoza è professore associato all'Università di Alicante e guida uno dei principali team di ricerca sull'OMR, responsabile dello sviluppo del sistema MuRET e dell'attuale stato dell'arte.

Il primo argomento di conversazione che affrontiamo è il contributo specifico alla tutela dei beni culturali in contesti biblioteconomici e museali. Stauder suggerisce un parallelo tra l'affermazione dell'HTR e l'invenzione del telescopio per gli astronomi o del microscopio per i biologi. Come questi strumenti hanno permesso ai ricercatori in quei campi di superare i loro limiti fisici nell'osservazione, così anche le scienze umane e sociali possono ora lavorare su quantità di informazioni altri-



menti impossibili da elaborare. Integrare in una biblioteca digitale un sistema HTR significa garantire la piena accessibilità e valorizzazione delle informazioni che custodisce. A volte, come Calvo-Zaragoza nota per il caso specifico degli archivi musicali, le istituzioni hanno speso, e continuano a spendere, risorse ingenti per conservare i supporti materiali, ma sono poco propense a investire nella conoscenza del loro contenuto. Elaborando il suo spunto potremmo definire queste situazioni di tutela come incompiute, perché conservano il bene materiale ma non valorizzano il bene immateriale del quale il supporto è latore. Mülberger invece presenta due casi pratici di utilizzo dell'HTR nei musei: l'informatizzazione dei diari di viaggio e dei resoconti delle esplorazioni, particolarmente utile per ricostruire la provenienza di oggetti e manufatti e accelerarne il processo di restituzione nel quadro della decolonizzazione dei musei europei, e la lettura delle etichette negli archivi dei musei naturali, per la quale è in fase di sperimentazione una soluzione robotica e informatica integrata. La parte successiva della discussione si sviluppa sui limiti attuali

della tecnologia e sulle prospettive future. Attualmente tra HTR e OMR vi è un'apparente asimmetria: l'HTR risulta più affidabile ed è più utilizzato dalla comunità scientifica di riferimento, mentre l'OMR è meno conosciuto, e sembra meno flessibile ai musicologi che vi si avvicinano come utilizzatori. Questo potrebbe indurre alla valutazione, erronea, che la tecnologia utilizzata per l'OMR sia meno evoluta di quella soggiacente all'HTR: il motore neurale alla base di entrambi è però sostanzialmente lo stesso. Quale allora la ragione di questa asimmetria apparente? Calvo-Zaragoza scompone il problema in tre aspetti distinti e correlati.

In primo luogo, l'oggetto di analisi dell'OMR è più complesso, non solo per la natura più articolata della scrittura musicale ma anche per la sua mutevolezza nel tempo. Per buona parte delle funzioni HTR è sufficiente il riconoscimento di un gruppo circoscritto di caratteri, che l'evoluzione della lingua nei secoli non ha modificato; la storia della musica occidentale è invece segnata da aggiornamenti e revisioni della semiografia musicale, per cui è difficile allenare un modello OMR unico che funzioni

diacronicamente per notazioni diverse.

Il secondo punto è la limitata consapevolezza da parte della comunità accademica interessata, che oscilla tra la diffidenza sulle possibilità applicative dell'OMR e il sottostimare il ruolo dell'annotazione dei dati. A volte, racconta Calvo-Zaragoza, chi contatta il suo gruppo di ricerca immagina di poter fornire una serie di pagine in una data notazione e di ottenere dal sistema direttamente una lettura codificata. La grande variabilità delle notazioni implica però che un modello allenato su manoscritti del settecento, ad esempio, non riuscirà in alcun modo a leggere un frammento di canto gregoriano. Questo rende ancora più cruciale la questione della conservazione e standardizzazione dei dati di allenamento, condivisa con l'HTR. Stauder auspica a questo riguardo la creazione di un quadro normativo che ne faciliti la condivisione e garantisca allo stesso tempo il rispetto della privacy e dei diritti di proprietà intellettuale.

Il terzo aspetto del problema è il finanziamento dei progetti. Solo l'avviamento di progetti di ampio respiro consente di produrre materiale annotato di dimensioni significative, porre all'attenzione della comunità accademica la questione della condivisione dei dati, e raggiungere i potenziali utilizzatori del software. L'integrazione, spesso più che altro nominale, dell'OMR in progetti Digital Humanities di piccole dimensioni rischia invece di frazionare la ricerca nel campo e creare modelli e database incompatibili. Il beneficio mag-

giore di un progetto di ampio respiro in questo settore è quello di segnare un riferimento per le esperienze successive. Nell'HTR è già accaduto: un progetto che prendesse avvio oggi difficilmente potrebbe ignorare gli standard posti da Transkribus, o eScriptorium, e dove decidesse di sviluppare un sistema proprietario faticherebbe a trovarne la giustificazione.

HTR e OMR, in quanto tecnologie la cui efficacia dipende dai dati e dalla loro gestione, hanno bisogno di uno sviluppo al contempo accentrato e cooperativo: accentrato per garantire uno standard di riferimento e non moltiplicare i costi di gestione della tecnologia, cooperativo per consentire a tutti i contributi di concorrere all'efficienza del sistema.

L'HTR di ultima generazione è costruito sull'architettura del transformer, che ha rimpiazzato i modelli precedenti basati sui Recurrent Neural Network (RNN). Mi incuriosisce il fatto che nell'OMR risultino tuttora più performanti: Calvo-Zaragoza spiega che questo accade per il loro intrinseco bias sequenziale – apprendono secondo una direzione, prima una parte del documento poi la successiva – utile per guidare la lettura di una pagina musicale. Nondimeno, architetture transformer-based potrebbero superare la performance dei RNN in presenza di una quantità sufficiente di dati d'allenamento, come è già accaduto in altri settori.

Parlando di Transformer non possiamo non menzionare GPT-4, superstar dell'AI che sta accentrando l'attenzione della

stampa, inaccessibile dall'Italia in questi giorni per sospetti sulla gestione della privacy. Domando se l'ultima creazione di OpenAI possa essere considerata il capostipite di una nuova generazione di grandi modelli linguistici (LLM) polifunzionali, capaci di competere in efficacia con sistemi specializzati di HTR e OMR. Al momento della discussione non sono ancora disponibili dati sperimentali, per cui le ipotesi si basano sulla performance di GPT-4 osservate in altri ambiti. Mülberger non ha dubbi sulle potenzialità di GPT-4, e prevede che presto incorporerà una funzione di riconoscimento del testo manoscritto base molto potente, per di più perfezionabile con allenamenti specifici. Osservo d'altro canto che molti dei compiti per i quali è utilizzato l'HTR oggi sono legati a scritture di nicchia per le quali è difficile trovare dati d'allenamento sul

web, che è il database di allenamento di GPT-4. Calvo-Zaragoza ricorda, a questo proposito, che se il web è pieno di parole, lo è un po' meno di musica. La percentuale di persone capaci di leggere la notazione musicale è molto minore di quella di persone alfabetizzate, e questo, mi dice scherzando, lo rassicura: se chat-gpt gli ruberà il lavoro, ci metterà almeno un anno o un paio. Per ora può stare tranquillo.

Dal mio punto di vista, non solo mi auguro questo non accada, ma ne dubito sinceramente. Il mio augurio è piuttosto che il mondo accademico e culturale colga l'occasione offerta da HTR e OMR per superare la visione, esatta ma limitata, dell'eredità testuale e musicale come oggetto di studio, valorizzandola invece in quanto patrimonio culturale condiviso.

NOTE

- 1 <https://www.theguardian.com/world/2023/feb/05/artificial-intelligence-uncovers-lost-work-bytitanof-spains-golden-age>
- 2 <https://www.ndr.de/geschichte/tagebuecher/Gefaelschte-Hitler-Tagebuecher-So-gefaehrlich-war-der-Fake,hitlertagebuecher108>.
- 3 [html https://neumz.com/reperitorium/](https://neumz.com/reperitorium/)
- 4 Lorenzo Casini. 2014. «“Noli me tangere”: i beni culturali tra materialità e immaterialità». *Aedon*, fasc. 1: 0-0. <https://doi.org/10.7390/76262>.

ABSTRACT

Data driven knowledge is revolutionizing many field of research, but handwritten cultural heritage doesn't seem to be able to benefit from this innovation. But what if AI itself could provide a

solution for the accessibility of documents? This article introduces the basic functioning and the latest developments of Handwritten Text and Music Recognition, Ai-based technologies that could provide archives and libraries, as well as museums, with a powerful tool to access on a scale their written content. We discuss the future perspectives with some of the leading experts in the field, focusing on the role this technology may play in connecting tangible documents and intangible cultural heritage.

PAROLE CHIAVE

HANDWRITTEN TEXT RECOGNITION; OPTICAL MUSIC RECOGNITION; AI; DEEP LEARNING; INTANGIBLE CULTURAL HERITAGE

AUTORE

ROBERTO RIBUOLI
ROBERTO.RIBUOLI@UNIROMA1.IT