



L'immagine immediata che abbiamo di internet è quella di uno sterminato archivio in continua crescita e di un luogo di incontro. Chi usa la rete reperisce informazioni e notizie, stabilisce e mantiene contatti con persone ed organizzazioni in quantità enormemente superiori a quelle realizzabili con le tecnologie che l'hanno preceduta.

Un sito di job placement come [SOUL](#), rilanciato dai motori di ricerca e dotato di semplici meccanismi di pubblicazione di opportunità di lavoro e raccolta curriculum, consente un incontro fra la domanda e l'offerta di lavoro in proporzioni neanche pensabili all'epoca.

Ancora negli anni '80 i neo-laureati dovevano scorrere pagine di annunci, inviare curriculum, presentazioni per posta e le aziende esaminare e confrontare centinaia di candidature. Come sfruttare adeguatamente le tecnologie informatiche per fare sì che le loro maggiori potenzialità non diventino un ostacolo?

Realizzare soluzioni banali con gli strumenti informatici è semplice. Filtri che selezionino le offerte di lavoro in base al settore di attività, alla figura professionale cercata, alla tipologia di contratto proposto, o che estraggano tra i candidati ad un'opportunità di lavoro quelli con una specifica laurea, o con qualche esperienza professionale nel settore, sono relativamente semplici da realizzare e sono disponibili non solo su SOUL.

I filtri attualmente utilizzati, tuttavia, non sono propriamente strumenti di valutazione, ma "setacci" a trama molto larga, senz'altro utili a ridurre il numero di offerte o candidature da valutare, ma non in grado di fornire precise risposte alle domande che un neo-laureato o un'azienda, si pongono: qual è l'offerta di lavoro migliore per me? Qual' è il candidato migliore tra quelli che si sono proposti?

È fondamentale osservare che la risposta ad una di queste domande non è un “assoluto invariante”, ma un “relativo” che muta nel tempo, al cambiare della situazione sociale ed economica. Le modalità d'incontro fra la domanda e l'offerta di lavoro, infatti, mutano rispetto alla “fase”. Se astrattamente cercassi il lavoro che ritengo ideale per me, potrei anche morire d'inedia prima di averlo trovato. Analogamente, un'azienda non può mancare alle proprie scadenze contrattuali o alle opportunità di mercato solo perché il candidato perfetto per svolgere uno specifico lavoro non è reperibile.

Vi sono tuttavia anche motivi più profondi, particolarmente in un sistema come SOUL orientato al job placement post-laurea, che rendono imprudente la ricerca di risposte assolute. Per esempio, un'attività lavorativa che non corrisponde ai miei desideri ideali potrebbe sviluppare mie potenzialità inespresse e darmi soddisfazioni che non sono al momento in grado di immaginare. Allo stesso modo, un candidato con un profilo non del tutto corrispondente a quello idealmente ricercato, potrebbe portare in azienda approcci nuovi ai problemi, aprendo opportunità non intraviste prima.

Una risposta efficace richiede un'intelligenza del quadro complessivo, storicizzata nel momento presente. Tradizionalmente, curriculum vitae ed offerte di lavoro sono composte da informazioni strutturate, rappresentate da campi con informazioni a scelta tra alternative prefissate e descrizioni discorsive in testo libero.

Le informazioni strutturate, come titoli di studio, facoltà, corso di laurea, voto conseguito, figura professionale, livello di familiarità con lingue straniere, che siano posseduti dal candidato o ricercati dall'azienda in una specifica offerta di lavoro, sono utili ai processi di filtraggio dei dati descritti sopra, ma possono anche essere utilizzate, in modo diverso, per fornire una prima base di valutazione.

Con una combinazione a coppie pesata delle corrispondenze tra i valori di tali elementi in un curriculum ed in un'offerta di lavoro, SOUL costruisce un indicatore, detto lineare, che dà una prima misura della prossimità tra il profilo di un candidato e quello ricercato dall'azienda. L'operazione è in realtà più complessa di quello che può sembrare a prima vista.

Ad esempio, le stesse competenze di base, se pur con approcci differenti ed in grado diverso, possono essere acquisite in corsi di laurea distinti, magari afferenti a diverse facoltà universitarie. Un laureato in Ingegneria Chimica nella facoltà di Ingegneria ed un laureato in Chimica nella facoltà di Scienze Matematiche Fisiche e Naturali avranno competenze di chimica

in comune. Al contrario un laureato in Matematica, nella stessa facoltà di Scienze Matematiche Fisiche e Naturali non avrà le stesse competenze in materia del collega chimico.

Se quindi un'offerta di lavoro richiede un candidato con una laurea afferente ad alcune facoltà universitarie, la corrispondenza con le caratteristiche dell'opportunità di lavoro deve essere pesata in base all'istogramma di frequenza dei settori disciplinari di riferimento nell'insieme di facoltà elencate, premiando così quegli elementi comuni all'insieme che, come verificato sperimentalmente, esprimono il profilo richiesto dall'azienda proponente l'opportunità di lavoro. Un discorso analogo vale qualora si voglia misurare la corrispondenza tra le aree di attività in cui il candidato ha fatto esperienza, o desidera farne, e quella proposta nell'offerta di lavoro.

L'indicatore lineare è indispensabile per la valutazione del grado di corrispondenza tra un curriculum ed un'offerta di lavoro, ma non è sufficiente. Limitandosi ai due esempi appena utilizzati l'offerta formativa universitaria è, da un canto, molto variegata: due ingegneri elettronici laureatisi nella stessa facoltà dello stesso ateneo, pur condividendo molte competenze di base, possono avere competenze specialistiche molto differenziate. D'altro canto, l'offerta formativa nell'ultimo decennio ha visto la nascita di nuovi corsi di laurea, caratterizzati da maggiore interdisciplinarietà del passato.

Analogamente, l'accelerarsi del progresso tecnologico e l'apparentemente inesauribile domanda di servizi innovativi da parte della società provocano la creazione di nuovi profili professionali, con un dinamismo che non può essere facilmente catturato in classificazioni rigide, per quanto rapidamente si riesca ad adattarle ed evolverle. Tutti gli elementi dinamici sia sul fronte della formazione universitaria che su quello delle professioni non possono essere agevolmente considerati dal semplice confronto fra campi "chiusi" e prefissati.

D'altronde, lo scopo delle componenti discorsive a testo libero di curriculum ed offerte di lavoro, mai tanto necessarie come al giorno d'oggi, è sempre stato quello di catturare ciò che non entra in rigidi schematismi: le competenze specialistiche acquisite dal candidato durante il lavoro di tesi o nelle esperienze di lavoro già realizzate, le capacità richieste e lo specifico contesto lavorativo e professionale per l'opportunità proposta dall'azienda. Anche i relativi testi, scritti (summary della tesi ad es.) peraltro nelle forme rituali tipiche delle specifiche attività descritte, devono essere tenuti in conto. Per poter valutare e confrontare un complesso così eterogeneo di elementi è sembrato naturale sfruttare tecniche che modellano l'attività dell'organo umano che meglio svolge tali valutazioni, il cervello.

Quando si sente l'espressione "reti neurali", chi ne ha già sentito parlare anche in forma divulgativa associa subito l'idea dei perceptron, uniti ad una singola uscita il cui valore è ricavato combinando molteplici ingressi, o delle reti multistrato di perceptron. Ciò avviene di solito per l'apparente somiglianza delle rappresentazioni grafiche di questi modelli con le strutture cellulari del tessuto nervoso. In realtà, i perceptron ed i loro derivati non si prestano bene al problema della valutazione della prossimità tra profili curricolari proposti e richiesti.

Innanzitutto, i modelli a perceptron sono dei classificatori, sono cioè utili a rispondere a domande come: questo oggetto è un contenitore di liquidi? E' un bicchiere, una tazza o una bottiglia? In secondo luogo, una rete a perceptron "impara" a risolvere una classe di problemi con una procedura di addestramento supervisionato, nella quale un insieme di dimensioni sufficienti di problemi campione della classe di interesse, dei quali è nota la soluzione, viene proposto iterativamente e ripetitivamente alla rete, e i parametri della stessa vengono progressivamente aggiustati in base alla correttezza o meno delle sue risposte, finché la rete non fornisce risposte giuste a tutti i problemi campione.

È evidente come tali modelli non si prestino al problema in esame. Intanto, al cambiare del quadro del mercato del lavoro, è necessario riaddestrare la rete, ma, cambiando lo scenario, cambierebbe, anche radicalmente, l'insieme di problemi campione, che dovrebbe ogni volta essere ricostruito con l'ausilio umano. L'onere non sarebbe sostenibile ed i risultati, probabilmente, incostanti, dato che la scelta di un candidato idoneo è un processo sicuramente più soggettivo di quello di decidere se un oggetto è una tazza, un bicchiere o altro.

Ma soprattutto, il problema di determinare quale offerta di lavoro è più adatta ad un certo profilo, o quale candidato è più adatto ad una determinata attività, è un problema di affinità contestuale, o meglio un problema di associazione tra dati, in base alla loro somiglianza. L'associazione di informazioni nuove con informazioni note è un'altra funzionalità in cui il cervello umano è estremamente efficiente, che si tratti di ricondurre una formula matematica mai vista prima a quelle che già conosciamo, o incontrando una persona mai vista prima, "riconoscere" in questa la figlia di due conoscenti che non vediamo da tempo. Ed è proprio questa la funzionalità che entra in campo quando è un essere umano che valuta la corrispondenza tra un curriculum ed un'offerta di lavoro.

Le Self-Organizing Maps (SOM) sono modelli utilizzati con successo da ormai trent'anni per riprodurre la funzionalità della memoria associativa tipica del cervello umano, ed il loro utilizzo in SOUL è la prima applicazione delle SOM per razionalizzare l'incontro fra la domanda e l'offerta nel mercato del lavoro.

Come dimostrato da decenni di tecniche di indagine medicale sempre più sofisticate, quando un dato sensorio, immagine, odore, sapore o suono, viene presentato al cervello umano, si attiva in risposta una porzione di corteccia cerebrale ben precisa, associata alla persona, cosa, pietanza, o parola corrispondente, se già nota, o alla più simile tra le tante che già conosciamo. Analogamente, la procedura di addestramento di una SOM distribuisce rappresentazioni di dati distinti su una griglia di celle, di solito bidimensionale come la corteccia cerebrale, in modo da riportare nella stessa cella, o in celle immediatamente adiacenti, i dati più somiglianti tra loro. L'addestramento non è supervisionato: il risultato non è noto a priori, ma dipende solo dall'insieme di dati utilizzati, che vengono distribuiti in modo equilibrato, così come la memoria associativa di un essere umano dipende dalle esperienze fatte e non da una classificazione assoluta a priori. La SOM così addestrata può essere successivamente utilizzata per determinare l'affinità di un nuovo dato a quelli memorizzati al suo interno.

Nel caso di SOUL, ogni singolo dato è costituito dal testo completo di un curriculum o da quello di un'offerta di lavoro. L'insieme delle offerte di lavoro attive negli ultimi mesi, che costituisce di fatto un' "immagine" della realtà attuale del mercato del lavoro, è utilizzato per l'addestramento della SOM, e per creare quindi la "mappa" alla quale è successivamente rapportato ciascun curriculum per verificare quali offerte di lavoro gli sono più vicine e quali più lontane.

Per misurare la "distanza" tra offerte di lavoro diverse e tra offerte e profili dei candidati, ogni documento, offerta di lavoro o curriculum, è descritto dalla frequenza delle parole che contiene. Sono state a questo fine utilizzate le principali tecniche di analisi statistica di documenti testuali sviluppate negli ultimi quindici anni. Ad esempio, congiunzioni, preposizioni, ed altri elementi grammaticali poco discriminanti vengono eliminati. Un termine che compare in tutti i documenti della base dati non può contribuire a distinguere un documento dall'altro e viene anch'esso eliminato.

Ancora, la frequenza delle parole contenute in un documento è normalizzata in base alla lunghezza del documento stesso e alla frequenza assoluta della parola, in tutta la base dati di SOUL. Ulteriori riduzioni su base statistica sono utilizzate per ridurre il significativo costo computazionale dell'addestramento.

L'indicatore neurale ricavato dalla SOM si rivela molto potente nel catturare e sfruttare il contenuto delle porzioni a testo libero. Facendo leva automaticamente sulle parole specialistiche o sigle del settore, consente di discriminare profili e competenze significativamente differenti nell'ambito della stessa attività professionale, si dimostra un mediatore capace nel raffronto tra una descrizione di un'attività lavorativa in termini generali e una descrizione della stessa in termini specifici, e anche in grado di evidenziare casi di "sovra

qualificazione" del profilo di un candidato rispetto all'attività proposta.

Come la memoria associativa umana, l'indicatore neurale basato sulla SOM ha i suoi difetti: déjà vu, somiglianze apparenti, dubbi. Un essere umano, incontrando una persona che gli sembra di conoscere, può risolvere il dubbio chiedendole come si chiama, ossia un'informazione strutturata. In modo del tutto analogo, SOUL impiega un mini sistema esperto per combinare il risultato dell'indicatore lineare con quello dell'indicatore neurale, raffrontandone certezze e dubbi, per estrarne una valutazione che sfrutti al meglio tutte le componenti dei dati inseriti da aziende e laureati.

Il raffronto dei due indicatori consente inoltre di fornire indicazioni analitiche sintetiche al candidato e all'azienda che lo valuta, aiutando la seconda in una scelta più oculata e il primo a migliorare la propria presentazione attraverso le informazioni riportate nel curriculum. Un'estesa campagna di test ha mostrato la validità della combinazione armonica dei due indicatori e come l'indicazione risultante, semplifichi i processi di valutazione di tutti gli attori coinvolti. Ovviamente, ci sono possibilità di miglioramento ulteriori, e si può ancora progredire su questa strada.

Si sa, l'intelligenza non è mai abbastanza...

- - -

Prof. Federico Massaioli

Coordinatore Gruppo HPC [CASPUR](#)

Consorzio Interuniversitario per le Applicazioni di Supercalcolo per Università e Ricerca, Roma