

atbv.it

Un algoritmo per accontentare tutti (o quasi) - ATBV

Temitope Ajileye

13-17 minuti

Di Temitope Ajileye

Ogni volta che uno studente universitario di matematica incontra una persona che studente universitario di matematica non è, la prima (o la seconda) domanda che quest'ultimo rivolge al primo è, immancabilmente: "Cosa fa il ricercatore di matematica?", con una intonazione che vuole dire: "Cos'altro si può dire delle quattro operazioni fondamentali?", oppure: "Esistono forse altre proprietà interessanti dell'elevamento a potenza?", e ancora: "L'analisi del liceo scientifico non è forse il punto d'arrivo della matematica?". Insomma, "Cosa ci potrà mai essere in matematica da scoprire?". Le risposte a questa domanda sono molteplici, una per ogni linea di ricerca seguita negli ultimi duecento anni.

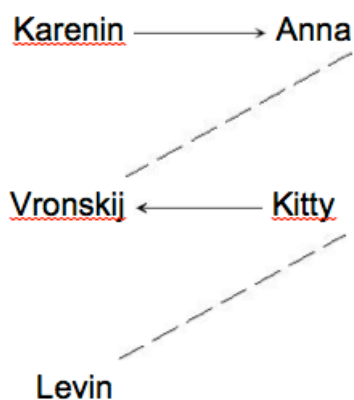
Tuttavia comunicare la matematica moderna a chi non si occupa di questa disciplina è difficile perché è necessario conoscere la lingua in cui viene prodotta. Senza questo, spiegare anche le proprietà più semplici diventa arduo; è come voler leggere Tolstoj in lingua originale senza conoscere il russo: bisogna dapprima imparare la lingua su frasi e testi più semplici. Per fortuna esistono alcune 'questioni pratiche' che interessano anche i matematici.

Molti sono i fenomeni, di cui facciamo esperienza quotidianamente, che hanno catturato la loro attenzione. Questi problemi sono facili da comunicare, non necessitano di un linguaggio complesso per poter essere trattati e le soluzioni sono talvolta eleganti. A volte, come nel caso che qui trattiamo, il solo fatto che esista una soluzione è sorprendente.

Un algoritmo a uso medico

Il problema

Il problema della stabilità di coppia è un problema importante ed è noto a tutti; per continuare il paragone con Tolstoj (e cominciare con tono leggero), analizziamo il problema nel suo romanzo *Anna Karenina*. I protagonisti di sesso maschile sono (per cognome) Levin, Vronskij, Karenin, mentre i protagonisti di sesso femminile sono Anna e Katerina detta Kitty, sua sorella. All'inizio del racconto Anna è moglie di Karenin mentre Kitty (che ha rifiutato Levin) e Vronskij sono “quasi” fidanzati. Le coppie sono come nella figura sottostante:



L'inghippo, e il romanzo, nasce quando Anna rende visita al fratello a Mosca e incontra Vronskij, che si infatua di lei fino al punto di seguirla nel viaggio di ritorno a San Pietroburgo. Le coppie potrebbero ancora reggere, senonché anche Anna

comincia a preferire Vronskij al marito. A questo punto ci sono due coppie in ognuna delle quali uno dei membri preferisce un membro dell'altra copia ed è ricambiato in questa preferenza. Le coppie non reggono: è una questione di matematica, oltre che di cuore.

Il matematico che legge *Anna Karenina* prova un desiderio incontenibile di trovare un metodo che possa evitare a Levin di passare dei mesi miserabili nella sua tenuta di campagna, a Karenin l'umiliazione pubblica arrecatagli dal tradimento di Anna, a costei la disperazione e l'ostracismo causati dall'essere sposata con un uomo e amarne un altro. Egli è spinto da motivazioni molto umane, ma rimane pur sempre un matematico. Per risolvere il problema, astrae i concetti di base dal contesto del romanzo e pensa, o forse grida, "Generalizziamo!".

Prendiamo due insiemi A e B, supponiamo contengano lo stesso numero di elementi, n . A ogni elemento x di A associamo una lista di 10 numeri, pescati, senza reintroduzione, da un'urna contenente gli interi dall'1 al 10 (il nome tecnico di questa operazione è permutazione). Questa è la lista delle preferenze di ogni elemento di A. Facciamo lo stesso per gli elementi di B. Un esempio per $n=3$ è il sottostante:

x_1	231	y_1	312
x_2	213	y_2	123
x_3	132	y_3	132

Dunque x_1 preferisce, nell'ordine, y_2 ; y_3 ; y_1 ; anche x_2 ha come prima scelta y_2 , ma la seconda e la terza sono diverse. y_2 invece ha come prima preferenza x_1 , mentre la seconda e la terza sono x_2 e x_3 .

Un abbinamento è una assegnazione di coppie formate da un elemento di A e un elemento di B. Per esempio:

$$x_2 + y_2 \quad x_1 + y_3 \quad x_3 + y_1$$

In questo esempio x_3 e y_1 sono la prima preferenza l'uno dell'altro, possiamo affermare con certezza che la coppia non scoppierà mai (diciamo che è stabile). Anche y_3 sta con la sua prima scelta ma il suo compagno, x_1 , guarda in modo piuttosto insistente verso y_2 , che preferisce a y_3 . y_2 è abbinato a x_2 ma ricambia le attenzioni di x_1 : queste due coppie non sono stabili. Definiamo ora in modo matematicamente preciso cosa intendiamo per stabilità, anche se dovrebbe essere già chiaro a questo punto.

Un abbinamento si dice **stabile** se, comunque prese due coppie $x+y$, $z+w$, con x e z nel primo insieme, y e w nel secondo insieme, se succede che x preferisce w a y , allora w non preferisce x a z .

Per alleggerire l'esposizione faremo uso della notazione $x : y_1 > y_2$ per dire che x preferisce y_1 a y_2 . Se modifichiamo l'abbinamento dato precedentemente, possiamo trovare:

$$x_1 + y_2 \quad x_2 + y_3 \quad x_3 + y_1$$

La prima e la terza coppia sono "perfette". La seconda coppia è formata da x_2 , che sta con la sua terza scelta, e y_3 , che sta anch'esso con la sua terza scelta.

Per quanto entrambi possano desiderare un'altra relazione, non potranno che stare assieme, perché tutti gli altri stanno con le loro prime scelte.

Sarebbe ancora facile sistemare le cose se le coppie da formare

fossero 3000 e non 3? Ma prima ancora, siamo sicuri che esista una soluzione, e che sia possibile evitare la formazione di coppie instabili? La risposta è sì, e la dimostrazione, trovata da David Gale a Lloyd Shapley, risale al 1962. Per il loro lavoro su questo problema e alcuni risultati collegati in ambito economico, Shapley venne insignito del premio Nobel per l'Economia nel 2012.

La dimostrazione è un bell'esempio di dimostrazione costruttiva, o algoritmica; ovvero sia non solo riusciamo a dire che la soluzione c'è, ma sappiamo anche trovarla! Per presentarla, sostituiamo il registro romantico a quello medico, ambito cui viene spesso applicato. Gli elementi di A ora sono neolaureati in medicina e gli elementi in B sono ospedali in cui cominceranno il tirocinio.

La soluzione

La ricerca della soluzione è articolata in turni, o iterazioni; ogni turno porta più vicino all'abbinamento stabile. Bisogna notare che la divisione in turni è solo funzionale a comprendere la soluzione, una volta che l'algoritmo è implementato in un computer, basta presentare la lista delle preferenze e l'abbinamento stabile viene calcolato in pochi secondi.

Nel primo turno i neomedici inviano il curriculum all'ospedale in cima alle loro preferenze. Alcuni ospedali potrebbero ricevere un solo curriculum, al quale rispondono con un "forse", altri ospedali potrebbero ricevere più proposte o nessuna. In ogni caso rispondono con "forse" al curriculum che preferiscono tra quelli ricevuti e "no" a tutti gli altri. Se un neomedico riceve un "forse", si considera temporaneamente impegnato e non manda altre proposte. Se in un turno successivo il "forse" si trasforma in "no", tornerà a inviare curriculum.

In ogni turno successivo ogni neomedico non impegnato invia un curriculum all'ospedale in cima alle sue preferenze, esclusi quelli che lo hanno già rifiutato, e ogni ospedale considera le proposte ricevute. Se uno dei nuovi curriculum è migliore di quello in forse, ammesso che ce ne fosse uno, rifiuta quello in forse e risponde "forse" al nuovo curriculum. In caso contrario rifiuta tutti i curriculum nuovi.

Dopo un numero finito di iterazioni arriva un turno in cui ogni curriculum è in "forse" e non ci sono domande nel turno successivo; il gioco è finito. A quel punto gli ospedali accettano la domanda che hanno in forse: l'abbinamento ottenuto è stabile. Cerchiamo di capire perché nel prossimo paragrafo.

Osservazioni

- Nel primo turno almeno un ospedale mette in forse una domanda, perché almeno un ospedale riceve dei curriculum. Può essere che la prima preferenza di ogni neomedico sia lo stesso ospedale.
- Per lo stesso motivo, in ogni turno almeno un ospedale che non aveva curriculum in forse risponde con "forse" a una delle proposte ricevute.
- Una volta che un ospedale mette un curriculum in forse, avrà sempre un curriculum in forse (non necessariamente lo stesso).
- Se gli abbinamenti da formare sono n , dopo n turni al massimo tutte le domande sono in forse e il matching è fatto.
- A ogni curriculum che il neomedico manda, il destinatario scende nella lista delle preferenze del medico.
- A ogni turno il curriculum in forse in ogni ospedale può rimanere invariato o salire nella lista delle preferenze dell'ospedale.

- Se alla fine del processo, consideriamo due copie $m_1 + H_1$ e $m_2 + H_2$, può accadere che $m_1 : H_2 > H_1$, ma questo vuol dire che m_1 ha inviato il curriculum ad H_2 prima di inviarlo ad H_1 e che H_2 l'ha rifiutato, oppure messo in forse e poi sostituito con m_2 , quindi $H_2 : m_2 > m_1$ e l'abbinamento è stabile.

n iterazioni per risolvere un problema “grande n ” è una stima molto buona; per molti dei problemi di computazione di solito ci vuole molto di più. Non è possibile dare una stima matematica migliore; si immagina il caso in cui i neomedici abbiano tutti la stessa identica lista di preferenze, perché magari alcuni ospedali sono o sono percepiti come nettamente migliori di altri. Dal punto di vista statistico, tuttavia, si possono ottenere stime migliori (per la maggior parte dei casi è sufficiente un numero basso e costante di iterazioni).

Questo non è l'unico modo di ottenere abbinamenti stabili, ma, per come è costruito, questo abbinamento è quello ottimale per chi fa le proposte, in questo caso i medici. Con ottimale vogliamo dire che nessun altro abbinamento stabile potrebbe assegnare a un dato medico un ospedale migliore di quello assegnatogli da questo abbinamento.

Capire perché il risultato fornito da questo algoritmo è quello ottimale per i proponenti è un poco più complicato; può essere un esercizio divertente, nei casi in cui n è piccolo, provare a trovare un abbinamento stabile migliore per i proponenti e, dopo aver constatato che ciò è impossibile, dimostrare che Gale-Shapley è davvero ottimale per i neomedici.

Applicazioni e sviluppi

Abbiamo dunque trovato un metodo per creare abbinamenti tra due insiemi che sia stabile e ottimale per i proponenti.

Nell'esempio utilizzato l'abbinamento è ottimale per i medici, ma basta invertire il processo (fare in modo che siano gli ospedali a mandare offerte) per ottenere un abbinamento che sia ottimale per gli ospedali. Questi non sono gli unici due algoritmi possibili, e la stabilità non è l'unica qualità che si può cercare in una coppia.

A questo punto dell'articolo ad alcuni potrà sembrare che tutto sommato non si tratti di una scoperta rilevante, o che le cose nel mondo "reale" vadano in maniera diversa. Niente di più sbagliato. Questo sistema è utilizzato, per esempio, dagli anni '50 negli Stati Uniti e abbina ogni anno 40.000 tirocinanti alle posizioni offerte dagli ospedali del paese. La cosa sorprendente è che già agli inizi il sistema utilizzato (senza dimostrazione) era in sostanza quello presentato da Shapley dieci anni dopo. L'istituto che se ne occupa è un'agenzia privata (National Resident Matching Program) e si può effettivamente trovare sul loro sito [una spiegazione dell'algoritmo](#) da loro utilizzato che assomiglia molto alla dimostrazione qui data.

Sempre sul loro sito, che consigliamo di visitare, si possono trovare [molti dati e statistiche sull'effettività del metodo](#). Questo sistema ha portato ordine in un "mercato" che stava diventando via via sempre più caotico (come accade oggi per l'ingresso nelle università blasonate degli USA). Prima che questo sistema prendesse piede, gli ospedali tendevano ad anticipare sempre più le offerte agli studenti di medicina, nella speranza di assicurarsi tutti i medici di cui avevano bisogno, con la conseguenza che queste offerte venivano ricevute da studenti che non avevano ancora deciso che specialità intraprendere.

Vari sistemi di abbinamento regionali sono stati utilizzati in Inghilterra, alcuni dei quali fallimentari. Le ricerche hanno trovato che quasi sempre i sistemi che funzionavano (cioè a cui gli studenti si affidavano) operavano cercando abbinamenti stabili, a dimostrazione che questo concetto ha davvero un'importanza oggettiva.

Per sviluppare ulteriormente questo problema, ci si può interrogare sulla [generazione delle preferenze](#); nell'esempio specifico, quali qualità i laureati cercano negli ospedali e viceversa, e come tali caratteristiche influiscano sulla qualità degli abbinamenti, qualità che, in termini pratici, chiamiamo soddisfazione lavorativa, buon funzionamento dell'ospedale, livello del servizio offerto al pubblico.

Non appena un fatto si dimostra vero, i matematici cercano sempre di estenderlo a domini vicini. Uno dei tentativi di estensione più importanti ha particolare interesse anche dal punto di vista "pratico". Capita infatti che molti studenti di medicina sposino altri studenti durante o dopo il corso degli studi; cercano dunque di entrare nello stesso ospedale. Esiste un algoritmo che riesca ad abbinare coppie agli ospedali in modo stabile? In questo caso ci sono due stabilità da bilanciare, quella degli ospedali e quella della coppia, la cui stabilità è minacciata da una eventuale separazione. La triste notizia è che modificare Gale-Shapley per abbinare coppie non funziona bene e questo ha portato molti studenti a operare fuori dallo schema NRMP, ma i matematici sono al lavoro e presto o tardi troveranno una soluzione!