

Complessità

Dall'acqua che bolle all'intelligenza artificiale

Giorgio Parisi

Il piano della lezione

- Che cosa vuol dire «complesso»: gli equilibri multipli.
- L'acqua che bolle e il ferro che si smagnetizza: le transizioni di fase.
- La simmetria nascosta: uno stato di equilibrio, due stati.
- I vetri di spin: un sistema con infiniti stati di equilibrio.
- Un trucco matematico, una soluzione «pazza» e il suo significato.
- Che cosa ci hanno insegnato i vetri di spin: il paesaggio accidentato.
- Dai vetri di spin alle memorie associative: il modello di Hopfield.
- Dalle memorie alle reti profonde e ai modelli linguistici.
- Perché tutto questo è fisica.

Che cosa vuol dire «complesso»?

La parola «complesso» sfugge di mano a chi cerca di darle una definizione precisa. A volte significa «complicato», cioè fatto di molti pezzi; altre volte «incomprensibile» o «imprevedibile».

La definizione che preferisco: un sistema complesso è un sistema il cui comportamento può cambiare nel tempo, cioè capace di mostrare molti comportamenti diversi senza perdere la propria identità. Un animale, quasi per definizione, è un sistema complesso: compie molte azioni diverse e passa rapidamente da uno stato all'altro.

Complesso non vuol dire complicato: un aereo è complicato, ma fa una cosa sola.

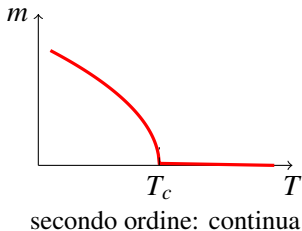
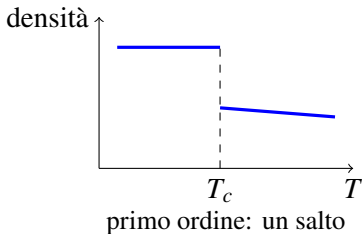
Il sistema complesso che conosciamo meglio è il nostro cervello. Prendiamo una sola delle sue capacità, la memoria: se decidiamo di ricordare qualcosa, possiamo restare su quel ricordo per cinque o dieci secondi, e anche di più, e poi passare rapidamente a ricordare qualcos'altro.

Comportamenti diversi corrispondono a stati di equilibrio diversi. Un sistema complesso ha molti stati di equilibrio.

L'acqua che bolle

A 100 °C l'acqua bolle: passa dallo stato liquido a quello aeriforme senza che cambi nulla nelle molecole, né nella loro forma né nelle loro interazioni. Cambia solo il modo in cui stanno insieme: è un fenomeno collettivo. A 0 °C diventa ghiaccio: le distanze tra le molecole sono quasi le stesse, ma i movimenti sono completamente diversi (il «rompete le righe» di una parata proiettato all'inverso).

Transizioni del primo ordine: le proprietà cambiano in modo discontinuo (a $-0,5$ °C l'acqua è solida, a $+0,5$ °C è liquida) e c'è un calore latente: la pentola resta a 100 °C finché non è evaporata tutta.



Il ferro che si smagnetizza

Pierre Curie, nella sua tesi di dottorato (1895): il ferro perde la capacità di essere attratto da una calamita se la temperatura supera i $756\text{ }^{\circ}\text{C}$ circa (il valore della tesi; oggi si misura $770\text{ }^{\circ}\text{C}$). Lo stesso avviene in tutti i ferromagneti: il nichel a poco meno di $400\text{ }^{\circ}\text{C}$, il cobalto a poco più di $1100\text{ }^{\circ}\text{C}$.

Qui non c'è nessun salto: la magnetizzazione diminuisce con continuità e si annulla alla temperatura critica. Niente calore latente. È una transizione del secondo ordine.

Pierre Weiss (1907): il ferro è fatto di magnetini elementari che a bassa temperatura si allineano con i vicini e producono un campo macroscopico; ad alta temperatura puntano in direzioni disperate e il campo risultante è praticamente nullo.

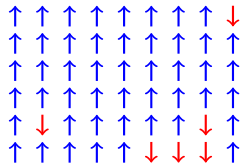
Il parametro d'ordine: una grandezza che è zero nella fase disordinata e diversa da zero in quella ordinata. Qui è la magnetizzazione.

Il modello di Ising (1925)

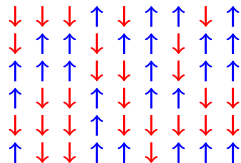
Spin $\sigma_i = \pm 1$ sui siti di un reticolo. Energia
 $H = -J \sum_{\langle ik \rangle} \sigma_i \sigma_k$: i primi vicini preferiscono essere
paralleli. Il peso di una configurazione è $e^{-H/T}$
(Boltzmann): $Z = \sum_{\{\sigma\}} e^{-H/T}$, $F = -T \log Z$.

Competizione tra energia, che vuole l'ordine, ed entropia,
che vuole il disordine: a bassa T vince la prima, ad alta T
la seconda.

Peierls (1936): in due dimensioni la transizione c'è
davvero. Onsager (1944): soluzione esatta in due
dimensioni, $T_c = 2,269 J$. In tre dimensioni nessuno l'ha
risolto, ma sappiamo quasi tutto con grande precisione.



$T < T_c$



$T > T_c$

Un modello «stupido» che contiene l'essenziale: i dettagli microscopici non contano.

La simmetria nascosta

L'energia non cambia se giriamo tutti gli spin: H è simmetrica per $\sigma \rightarrow -\sigma$.

Sopra T_c lo stato di equilibrio rispetta la simmetria: $m = 0$.

Sotto T_c ci sono due stati di equilibrio, $m = +m_0$ e $m = -m_0$, uno l'immagine speculare dell'altro, e il sistema ne sceglie uno (per un campo esterno infinitesimo, o per la sua storia). La simmetria non è sparita: è nascosta, e si vede nell'esistenza dell'altro stato. **Rottura spontanea di simmetria.**

Lo stesso schema si ritrova dappertutto: nei cristalli (la simmetria di traslazione), nei superconduttori, nel bosone di Higgs.

Vicino a T_c il sistema fluttua su tutte le scale (invarianza di scala) e le grandezze vanno come potenze di $T - T_c$, con esponenti universali: l'acqua al punto critico e il ferro a T_c hanno gli stessi esponenti. Kenneth Wilson (1971, premio Nobel 1982) ha spiegato perché con il gruppo di rinormalizzazione.

Fin qui: uno stato di equilibrio, o due stati legati da una simmetria. Possiamo avere sistemi con moltissimi stati?

Equilibri multipli?

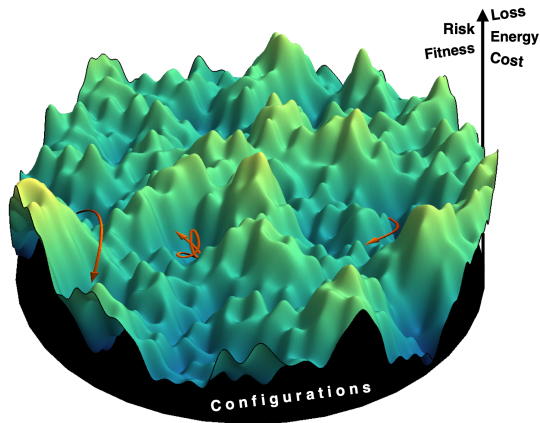
Un sistema semplice, come quelli storicamente studiati dalla fisica, ha un solo comportamento o pochissimi comportamenti diversi: l'acqua ghiaccia o evapora, niente di più; un metallo è magnetico o non lo è, conduce o non conduce.

Comportamenti diversi corrispondono a stati di equilibrio diversi, a fasi diverse. Fissati temperatura e pressione, un materiale sta di solito in una sola fase; a valori particolari due fasi coesistono (è la transizione), tre fasi solo in un punto triplo, mai di più: la regola delle fasi di Gibbs.

Senza aggiustare finemente i parametri, possiamo costruire sistemi che mostrano un gran numero di fasi diverse? Apparentemente no.

I vetri di spin hanno dimostrato che esistono sistemi fisici che, nel limite di volume infinito, hanno un numero infinito di fasi.

Il paesaggio accidentato



Pensiamo all'energia come a un paesaggio sopra lo spazio delle configurazioni. Per il ferro: due valli. Per un sistema complesso: moltissime valli separate da montagne, di profondità diverse, le piccole dentro le grandi.

Goldstein (1969), per i vetri: a bassa temperatura un liquido sottoraffreddato esplora lo spazio delle fasi saltando tra minimi amorfi diversi, separati da barriere di energia potenziale.

Debreu (1970), economista, premio Nobel: *Economies with a finite set of equilibria*. Per gli economisti gli equilibri multipli erano già familiari.

La domanda per un fisico: esiste un sistema con un'energia scritta in una riga che abbia questo paesaggio? E si riesce a calcolare qualcosa?

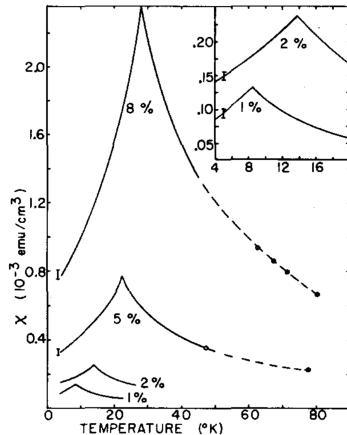
I vetri di spin metallici

Verso gli anni Sessanta si studiano leghe con una piccola percentuale di ferro diluita nell'oro, per capire il magnetismo dei singoli atomi di ferro (l'effetto Kondo).

Sorpresa: una transizione di fase che non sparisce quando la concentrazione tende a zero. La temperatura di transizione è proporzionale alla concentrazione (Cannella e Mydosh, 1971).

Gli atomi di ferro occupano posizioni casuali. L'interazione tra due spin, mediata dagli elettroni dell'oro, oscilla con la distanza: $J(r) \propto \cos(k_F r)/r^3$ (Ruderman, Kittel, Kasuya, Yosida). Metà delle coppie vuole spin paralleli, metà antiparalleli: nessuno sa quali.

Sistemi disordinati: ogni campione ha le sue interazioni, diverse da quelle di tutti gli altri.



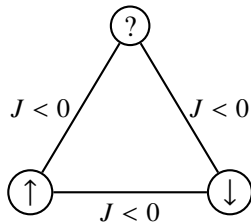
La frustrazione

Tre spin, e tutte e tre le coppie vogliono essere antiparallele: impossibile accontentare tutti. Una coppia resta scontenta, e possiamo scegliere quale: lo stato di energia minima non è unico.

In un vetro di spin ci sono miliardi di triangoli frustrati, intrecciati tra loro: moltissimi modi quasi equivalenti di scontentare il minor numero possibile di coppie.

In parole: cento persone che si conoscono tra loro. Ad alcune piace la matematica, ad altre la letteratura; ad alcuni gli scacchi, ad altri il poker; metà laziali e metà romanisti ... Le persone con gli stessi interessi amano stare insieme, ma gli argomenti sono così tanti che alla fine la simpatia di ogni coppia è una variabile casuale.

Ogni modo di dividerle in gruppi lascia molti scontenti, e ci sono moltissime divisioni diverse, tutte ugualmente imperfette.



Il modello di Edwards e Anderson (1975)

Un modello di una riga che cattura quasi tutto:

$$H = -\frac{1}{2} \sum_{\langle ik \rangle} J_{ik} \sigma_i \sigma_k ,$$

dove le J_{ik} tra primi vicini sono variabili casuali indipendenti, con media zero e $\overline{J_{ik}^2} = 1$.

C'è una magnetizzazione spontanea? Sì, ma ogni spin ha la sua: $m_i = \langle \sigma_i \rangle$, con segno diverso da spin a spin, e la somma di tutte è zero. Il vecchio parametro d'ordine non serve.

La grande idea: il parametro d'ordine è

$$q_{EA} = \frac{1}{N} \sum_i m_i^2 ,$$

zero sopra T_c e positivo sotto. La suscettività $\chi = (1 - q_{EA})/T$ ha una derivata discontinua a T_c : la cuspidè dell'esperimento.

Il modello di Sherrington e Kirkpatrick (1975) e il trucco delle repliche

Tutti gli spin interagiscono con tutti, J_{ik} casuali con $\overline{J_{ik}^2} = 1/N$: un modello di campo medio, che deve essere risolubile esattamente.

Si vuole la media sul disordine di $F = -T \log Z$. La media di Z^n è facile per n intero: n copie del sistema, le «repliche», con le stesse J . Poi si manda n a zero:

$$\overline{\log Z} = \lim_{n \rightarrow 0} \frac{\overline{Z^n} - 1}{n}.$$

Un trucco spregiudicato (continuare dagli interi ai reali), ma non nuovo: lo faceva già Regge.

Nel calcolo compare una matrice Q_{ab} ($a, b = 1, \dots, n$) di sovrapposizioni tra repliche. Sherrington e Kirkpatrick fecero l'ipotesi, dall'aria innocente, $Q_{ab} = q$ per ogni $a \neq b$: tutte le repliche sono equivalenti (simmetria delle repliche).

Risultato: formule semplici, ma l'entropia a temperatura zero è negativa, $-1/(2\pi)$. Qualcosa non va. De Almeida e Thouless (1978): la soluzione è instabile a bassa temperatura.

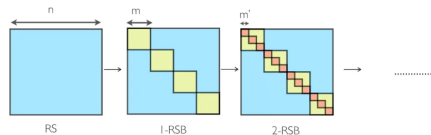
La soluzione (1979)

Si provò a dividere le n repliche in n/m gruppi di m , con m intero (Bray e Moore 1977, Blandin 1978, De Dominicis e Garel 1979): energia dello stato fondamentale $-0,789$ contro il $-0,76 \pm 0,01$ delle simulazioni. Non andava.

Nel 1979, venendo da tutt'altro (la cromodinamica), feci l'ipotesi azzardata che m potesse essere un numero non intero tra 0 e 1: per esempio n repliche divise in $2n$ gruppi di mezza replica ciascuno. Energia e calore specifico giusti, entropia $-0,01$: ancora negativa, ma quasi zero.

Poi: dividere i gruppi in sottogruppi, e così via, infinite volte. Al limite la matrice è descritta da una funzione $q(x)$, $0 \leq x \leq 1$. Entropia zero, energia $-0,7633$.

Il referee: *L'approccio non ha senso, ma i numeri che si ottengono dalle formule sono ragionevoli, quindi si può pubblicare. L'ultima osservazione non vale la carta su cui è scritta e va tolta.*

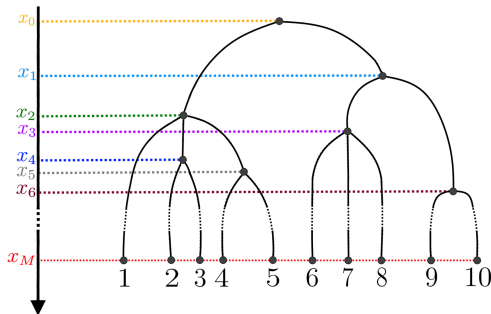


Che cosa vuol dire? Infiniti stati, organizzati ad albero (1982-1984)

Per tre anni non ho saputo che cosa significasse $q(x)$. Poi, con Mézard, Soursas, Toulouse e Virasoro: un vetro di spin ha molti stati di equilibrio α , ciascuno con il suo peso w_α e le sue magnetizzazioni m_i^α ; la somiglianza tra due stati è $q_{\alpha\gamma} = \frac{1}{N} \sum_i m_i^\alpha m_i^\gamma$, e la $q(x)$ codifica la distribuzione di probabilità di queste sovrapposizioni.

Gli stati sono organizzati come le foglie di un albero, e la distanza tra due stati è la distanza lungo l'albero: ultrametricità. Rami a ogni livello, infiniti rami in ogni intervallo (la costruzione matematica: Ruelle, poi Aizenman e Bolthausen).

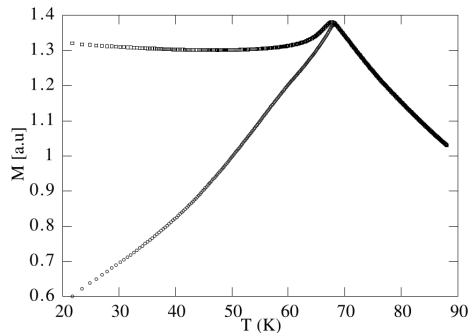
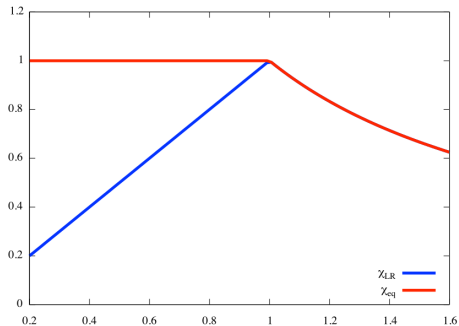
Una classificazione gerarchica, come la tassonomia degli esseri viventi; ma qui non viene dalla storia: è una proprietà dell'equilibrio.



Una predizione chiara, e l'esperimento

Se applichiamo un piccolo campo a bassa temperatura, il sistema resta nel suo stato:

$\chi_{LR} = (1 - q_{EA})/T$. Se applichiamo il campo ad alta temperatura e poi raffreddiamo, il sistema sceglie lo stato migliore: $\chi_{FC} = \frac{1}{T} \int_0^1 dx (1 - q(x))$. Se $q(x)$ non è costante, le due suscettività sono diverse.



A sinistra la teoria di campo medio; a destra l'esperimento (Djurberg, Jonason e Nordblad). **Negli esperimenti sono davvero diverse.**

I matematici vengono in soccorso

La forma «pazza» della matrice Q equivale a questa distribuzione di probabilità molto complessa. Ero molto scettico sulla possibilità di dimostrare che l'ipotesi fosse corretta.

Alla fine degli anni Novanta Francesco Guerra introduce idee nuove: nel 2003 dimostra che la formula è un limite superiore per l'energia libera. Michel Talagrand (2006) dimostra che è esatta. Dmitry Panchenko (2013) dimostra l'ultrametricità.

Oggi la dimostrazione è compatta: pochi passi indipendenti, ciascuno dimostrabile per conto suo. Il lavoro dei matematici è stato fondamentale per chiarire le ipotesi fisiche che stanno sotto le repliche e per togliere ogni dubbio sulla correttezza dei risultati.

Però nessuno è ancora riuscito a trasformare le repliche stesse in una dimostrazione rigorosa. Ho qualche idea, ma trasformare idee vaghe in un teorema non è facile.

La cornucopia dei vetri di spin

P.W. Anderson (1989): *Per me il risultato chiave è la bella rivelazione della struttura del «paesaggio accidentato» casuale che sta sotto molti problemi complessi di ottimizzazione. (...) Alcuni hanno sminuito l'importanza della struttura ultrametrica, dicendo che è una proprietà del modello SK e non dei vetri di spin fisici. Questa critica manca il punto: i vetri di spin fisici e il modello SK sono solo il punto di partenza di una sorprendente cornucopia di applicazioni dello stesso modo di pensare. Anderson aveva ragione.*

Dentro la fisica: la dinamica lenta e l'invecchiamento dei vetri; i vetri strutturali; le sfere dure e il jamming; polimeri e interfacce in mezzi disordinati.

Fuori dalla fisica: l'ottimizzazione (il commesso viaggiatore, l'assegnazione); la soddisfacibilità (K -SAT); i codici a correzione d'errore; il compressed sensing; le reti neurali.

In tutti questi casi lo spazio delle soluzioni ha la struttura dei vetri di spin: molte valli, organizzate gerarchicamente. Saperlo ha permesso di costruire algoritmi nuovi.

Una transizione di fase in informatica

N variabili vere o false, M vincoli, ciascuno su tre variabili (« x_1 oppure non x_5 oppure x_9 »). Esiste un'assegnazione che li soddisfa tutti? È il problema NP-completo per eccellenza (3-SAT).

Per N grande conta solo il rapporto $\alpha = M/N$. Con Mézard e Zecchina (2002): per $\alpha < 4,267$ una soluzione esiste quasi certamente, per $\alpha > 4,267$ quasi certamente no. Già prima della soglia, intorno a $\alpha \approx 3,86$, le soluzioni si rompono in grappoli lontani tra loro: è il paesaggio dei vetri di spin, ed è lì che gli algoritmi si bloccano.

L'analisi ha suggerito un algoritmo nuovo (*survey propagation*) che risolve istanze con un milione di variabili vicino alla soglia.

I matematici hanno dimostrato dopo una dozzina d'anni che la soglia è quella giusta per vincoli con molte variabili (Ding, Sly e Sun, 2015); per tre variabili la dimostrazione manca ancora.

Complessivamente hanno fatto un buon lavoro.

I vetri veri e le sfere dure

Il vetro delle finestre è un liquido che non è riuscito a cristallizzare: raffreddandolo la viscosità cresce di più di dieci ordini di grandezza in un intervallo modesto di temperatura, senza che la struttura cambi in modo visibile.

Il modello più semplice: sfere dure, $U = \infty$ se si compenetrano e $U = 0$ altrimenti. A bassa densità un fluido; aumentando la densità la viscosità esplode e la diffusione si ferma (nelle simulazioni cresce di un fattore 10^6 , poi non si misura più). A pressione infinita le sfere si toccano e non possono muoversi: il jamming.

Nel limite di dimensione spaziale infinita il problema si risolve esattamente con lo stesso formalismo dei vetri di spin (con Charbonneau, Kurchan, Urbani e Zamponi, 2014). Compaiono esponenti irrazionali: $\gamma = 0,41269 \dots$ e $\kappa = 1,41574 \dots$. Le simulazioni in tre dimensioni danno 0,4 e 1,4.

La rottura della simmetria delle repliche c'è anche qui, in un sistema fatto solo di palline che si urtano.

Dai vetri di spin ai neuroni: la memoria associativa

- Memoria del computer (indirizzabile): a ogni indirizzo troviamo delle informazioni. Come una cassettera.
- **Memoria associativa:** abbiamo una conoscenza parziale di un elemento e richiamiamo alla mente l'intero elemento (vediamo il volto e ci ricordiamo la persona: il nome, dove abita, che lavoro fa).

Si può costruire una memoria associativa usando come supporto una memoria indirizzabile: si scorrono tutti gli elementi. Il cervello non funziona così: risponde in pochi cicli (un neurone ha un tempo tipico di 40 ms, siamo macchine a 25 hertz) e il tempo di risposta dipende pochissimo dalla lunghezza dell'elenco.

L'album delle figurine: «ce l'ho, ce l'ho, mi manca», alla stessa velocità con dieci figurine o con mille.

Come costruire con neuroni digitali un sistema che funzioni da memoria associativa?

Il modello di Hopfield (1982)

John Hopfield, PNAS, primo aprile 1982: le proprietà computazionali utili agli organismi possono emergere come proprietà collettive di sistemi con un gran numero di componenti semplici.

Un neurone è attivo o quiescente: uno spin $\sigma_i = \pm 1$. Le sinapsi sono le J_{ik} , simmetriche; la dinamica fa scendere l'energia

$$H = -\frac{1}{2} \sum_{ik} J_{ik} \sigma_i \sigma_k ,$$

e gli stati di equilibrio (i minimi) sono i ricordi. Per memorizzare M configurazioni ξ^ν ($\nu = 1, \dots, M$):

$$J_{ik} = \frac{1}{N} \sum_{\nu=1}^M \xi_i^\nu \xi_k^\nu \quad (\text{regola di Hebb: chi si accende insieme si lega insieme}).$$

Mostrando al sistema una parte di un ricordo, la dinamica completa il resto.

È un modello di Sherrington e Kirkpatrick con le interazioni scelte da noi: Hopfield stesso cita i vetri di spin. Un piccolo passo indietro (le sinapsi simmetriche non sono realistiche) e un gigantesco passo avanti: si può usare tutta la meccanica statistica dell'equilibrio.

Il modello di Hopfield in parole

Abbiamo 100 persone: si conoscono tra loro.

Ad alcune piace la matematica, ad altre la letteratura. Ad alcuni piace giocare a scacchi, ad altri a poker, metà laziali e metà romanisti ...

Le persone interessate allo stesso argomento amano stare insieme.

Come trovare vicine tutte le persone a cui piace la matematica? Si selezionano alcune persone a cui piace la matematica e si dice ad alta voce: *Stiamo formando un gruppo, chi vuole venire con queste persone?* La maggior parte delle persone interessate alla matematica si unisce.

Persone $\rightarrow$ neuroni. Amicizia $\rightarrow$ forza sinaptica. Argomenti $\rightarrow$ stati di equilibrio.

La formulazione matematica di questo problema è il modello di Hopfield. Nel vetro di spin gli argomenti sono così tanti che nessuno li ha scelti: gli stati di equilibrio ci sono lo stesso, ma non sappiamo quali.

Quanti ricordi?

Se proviamo a memorizzare troppe cose, andiamo in uno stato di confusione totale, dove non ricordiamo più niente. Quale è la capacità della memoria?

Amit, Gutfreund e Sompolinsky (1985), con le tecniche dei vetri di spin: con la regola di Hebb si possono memorizzare fino a $M = \alpha_c N$ configurazioni, $\alpha_c \approx 0,14$; una sola in più e crolla tutto. È una transizione di fase.

Elizabeth Gardner (1988): scegliendo le sinapsi nel modo migliore si arriva a circa $2N$. In tutti i modelli la capacità è dell'ordine del numero di sinapsi: circa un bit per sinapsi.

Il nostro cervello ha 10^{14} sinapsi: 10 terabyte, qualcosa come dieci milioni di libri senza figure.

1985: una scuola invernale a Les Houches con fisici, lo stesso Hopfield, biologi e informatici (tra cui Yann LeCun). **Le reti neurali moderne nascono con un contributo essenziale dei fisici, che pongono domande precise nel limite di infiniti neuroni:** quanti stati di equilibrio, quale percentuale di errori, come dipendono dai parametri del sistema.

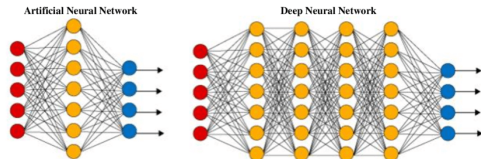
Dalla memoria all'apprendimento

Hopfield risolve la memoria. Il problema successivo è imparare dagli esempi: dato un input (una fotografia) produrre l'output giusto (ponte o diga).

Hinton, Ackley e Sejnowski (1985), la macchina di Boltzmann: un modello SK con neuroni di input (bloccati), di output e nascosti; le sinapsi sono scelte in modo che, con la distribuzione di Boltzmann, l'output sia quello voluto. Funziona, ma è lenta.

Rumelhart, Hinton e Williams (1986): reti a strati e backpropagation. Si scende lungo il gradiente dell'errore, e ogni sinapsi viene a sapere quanto dell'errore è colpa sua.

2012: Krizhevsky, Sutskever e Hinton, AlexNet, 60 milioni di sinapsi, addestrata su due schede grafiche: errore del 15% contro il 26% del secondo nel riconoscimento di immagini (ImageNet). Da allora reti profonde ovunque, con centinaia di miliardi di sinapsi.



E la teoria?

Uno strato (il perceptron): ottima teoria, fatta dai fisici negli anni Ottanta e Novanta con le repliche.

Due strati: buona teoria.

Molti strati: poca teoria. Possiamo trovare buoni progetti per tentativi ed errori, ma non sappiamo come la qualità delle risposte dipenda dal numero di neuroni e di strati, né quali caratteristiche rendano un problema facile o difficile da imparare.

La situazione ricorda le macchine a vapore prima della termodinamica: funzionavano (Newcomen 1712, Watt 1769), ma nessuno sapeva calcolarne il rendimento finché, verso il 1850, non si capì che il calore è una forma di energia.

Le macchine ci sono, la termodinamica non ancora. Dagli anni Novanta fisica e intelligenza artificiale si erano allontanate; negli ultimi anni si stanno riavvicinando, e cominciano ad arrivare risultati teorici anche sui transformer.

I grandi modelli linguistici

A questi modelli viene presentata un'incredibile quantità di testi (più di diecimila miliardi di parole) e vengono addestrati a indovinare la parola successiva. L'architettura è il transformer (2017): una rete che decide a quali parole precedenti prestare attenzione («Ho visto un cappuccino»: se prima c'era «bar» o «convento» l'ambiguità sparisce). Centinaia di miliardi di sinapsi, i più grandi oltre mille miliardi; il cervello ne ha 10^{14} .

Funzionano in modo sorprendente: rileggono e riscrivono testi, traducono, scrivono programmi, e le ultime versioni fanno una specie di ragionamento.

Un limite essenziale: conoscono solo il mondo delle parole. Sono «parolai», disincarnati: non hanno un corpo e non hanno un modello interno del mondo. Noi e gli altri animali ne abbiamo uno, precedente e indipendente dal linguaggio. Un modello linguistico è inutile per guidare un'automobile.

Si dice che certe capacità «emergano» quando il modello è abbastanza grande, ma nessuno sa caratterizzarle con precisione. È un problema di fisica dei sistemi complessi, e la teoria non c'è.

Perché il premio Nobel per la fisica a Hopfield e Hinton

Ho scritto a un amico qualche anno fa:

Sono assolutamente convinto che un premio Nobel per «le scoperte rivoluzionarie e lo sviluppo delle reti neurali artificiali che costituiscono le pietre miliari dell'intelligenza artificiale» sia una necessità. Penso che il Nobel dovrebbe continuare a espandersi in più regioni della conoscenza: la fisica sta diventando sempre più ampia e comprende molte aree di conoscenza che in passato non esistevano o non facevano parte della fisica.

Questa espansione era stata prevista da P.W. Anderson nel 1972, in *More is different*:

Così, con una complicazione crescente a ogni stadio, saliamo la gerarchia delle scienze. Ci aspettiamo di incontrare a ogni stadio domande affascinanti e, credo, molto fondamentali. (...) Certamente ci sono più livelli di organizzazione tra l'etologia umana e il DNA che tra il DNA e l'elettrodinamica quantistica, e ogni livello può richiedere una struttura concettuale interamente nuova.

Per concludere

Dall'acqua che bolle all'intelligenza artificiale il filo è uno solo: tanti elementi semplici che interagiscono, e un comportamento collettivo che non si legge nei singoli.

- Uno stato di equilibrio: l'acqua, il vapore.
- Due stati, legati da una simmetria nascosta: il ferro sotto T_c .
- Infiniti stati, organizzati ad albero: i vetri di spin.
- Gli stati che scegliamo noi: la memoria di Hopfield.
- Gli stati che la macchina impara da sola: le reti profonde, i modelli linguistici.

A ogni passo sono serviti concetti nuovi (il parametro d'ordine, la rottura di simmetria, le repliche, la $q(x)$), e a ogni passo la matematica rigorosa è arrivata dopo la fisica, a volte vent'anni dopo.

Dei sistemi complessi sappiamo calcolare i casi più semplici. Dei più importanti, il cervello e le macchine che abbiamo costruito, non abbiamo ancora la termodinamica. C'è molto lavoro per chi comincia adesso.

Il potere economico

Oggi alcuni usano ChatGPT, altri Gemini, Claude, DeepSeek. C'è però la possibilità che tutto questo evolva verso un monopolio, come è accaduto con i motori di ricerca: negli anni Novanta ce n'erano decine, oggi uno solo fa circa il 90% delle ricerche.

Il monopolio è lo scenario più terrificante: uno solo controlla l'accesso generale all'informazione. Un oligopolio non sarebbe molto diverso, specialmente se gli attori vengono dalla stessa parte del mondo e hanno la stessa ideologia.

I costi (decine di migliaia di GPU, centri di calcolo da miliardi di dollari) selezionano pochissimi attori, che accumulano dati, potenza di calcolo e profitti.

Il lavoro: Geoffrey Hinton (settembre 2025) prevede che l'IA renderà poche persone molto più ricche e la maggioranza più povera; non è colpa della tecnologia, ma del sistema economico. Le risposte sono politiche: un reddito di base, la riduzione dell'orario di lavoro («lavorare meno, lavorare tutti»).

Il potere politico

Chi controlla i modelli controlla quello che miliardi di persone leggono, imparano e credono: l'IA sta soppiantando siti e fonti, e le risposte arrivano da una sola voce.

Lewis Carroll, *Attraverso lo specchio* (1871): «Quando io uso una parola», disse Humpty Dumpty, «significa esattamente quello che decido io, né più né meno». «La questione è», disse Alice, «se lei può fare in modo che le parole significhino tante cose diverse». «La questione è», rispose Humpty Dumpty, «chi è che comanda. Tutto qui».

Chi decide che cosa un modello può dire e che cosa no? Oggi poche aziende private, quasi tutte in due Paesi, senza nessun controllo pubblico.

L'IA è diventata un fattore della competizione tra grandi potenze: controlli sulle esportazioni, corsa ai chip, applicazioni militari. Bisogna incominciare a riflettere su come smorzare queste tensioni prima che si trasformino in qualcosa di peggio.

Delegare il pensiero alle macchine

Vannevar Bush, ottant'anni fa: *Non c'è nessun surrogato per il pensiero maturo. Ma il pensiero creativo e il pensiero essenzialmente ripetitivo sono due cose molto diverse. Per il secondo ci sono, e ci possono essere, potenti ausili meccanici.*

Delegare il pensiero ripetitivo va bene. Delegare il giudizio no: una capacità che non si esercita si perde.

A scuola: i ragazzi chiedono all'IA il tema «scritto con un linguaggio da ragazzi e con qualche errore». I compiti a casa sono diventati quasi obsoleti; bisogna insegnare a usare l'IA in modo critico, in un mondo in cui gli adulti stessi fanno fatica a orientarsi.

Nella ricerca: usare la macchina per esplorare, calcolare, controllare, leggere più in fretta. Non delegare la scelta del problema; rifare da soli i calcoli semplici; un passaggio che non si sa spiegare non deve restare nel proprio articolo. È lecito usare uno strumento che non si capisce fino in fondo; non è lecito nascondere che non lo si capisce.

L'IA deve essere usata come uno strumento, non come un'autorità.

Il futuro della Matematica

La settimana scorsa sono stati rilasciati da Open-AI più di 700 teoremi.

- Un quasi dimostrazione della congettura di Riemann.
- La prova della congettura di Mézard Parisi su vetri diluiti.
- La prova della formula di Franz Parisi per il perceptrone sferico.

Molti altre prove verranno rilasciate nel futuro.

Chi sarà disposto a studiare un problema per un anno se poi c'è il rischio di venir bruciati sul filo del rasoio dall'IA?

Forse la matematica fatta da soli uomini è destinata a diminuire sempre di più e verrà sostituita da quella fatta da una collaborazione uomo macchina (o dalle sole macchine?)

I pericoli per l'umanità

- Le armi autonome letali (droni, robot killer) scelgono e attaccano i bersagli da sole: una nuova corsa agli armamenti. **La scelta di chi uccidere non può essere lasciata a un algoritmo.** Serve un bando, come per le armi chimiche e biologiche.
- Sorveglianza di massa, *social scoring*, manipolazione: strumenti di dominio nelle mani di governi e di imprese.
- Sistemi progettati per generare dipendenza, illusione, perdita di autonomia (le chat che diventano amici, consiglieri, psicologi).
- Decisioni di vita o di morte (guerra, ordine pubblico, frontiere, sanità, giustizia): mai a un sistema di IA.
- La superintelligenza: il suo sviluppo non va autorizzato finché non c'è un ampio consenso scientifico sulla sicurezza e un chiaro consenso pubblico.

Dall'appello presentato a papa Leone XIV e all'Assemblea generale dell'ONU (settembre 2025). Solo gli esseri umani sono soggetti morali e giuridici: i sistemi di IA sono e devono restare oggetti.

Che fare: un grande centro pubblico di ricerca sull'IA

Servono dati, modelli e programmi aperti, costruiti nell'interesse pubblico. Possono venire solo da un grande centro di ricerca pubblico, che in questo momento storico è immaginabile solo in Europa.

Un precedente: Linux. La comunità scientifica e informatica si è ribellata ai sistemi proprietari e chiusi e ha costruito un'alternativa aperta, che oggi fa funzionare la maggior parte dei server del mondo e tutti i supercalcolatori più potenti.

Il Manifesto per un Centro europeo di ricerca sull'IA (Accademia delle Scienze di Bologna): ricerca di base per capire i meccanismi dell'IA, come la termodinamica ha fatto con la macchina a vapore; il capitale umano al primo posto; modelli aperti e accessibili; indipendenza scientifica; efficienza energetica; etica fin dall'inizio.

Il presidente Mattarella (19 dicembre 2025): «I Paesi dell'Unione dispongono, insieme, delle risorse necessarie per realizzare un modello di intelligenza artificiale originale, trasparente, sicuro, attento ai diritti».