Predizione e contraddizione dalla biologia all'IA

PARTE SECONDA Intelligenza artificiale open-ended

Francesco Bianchini

Dipartimento di Filosofia e Comunicazione

francesco bianchini5@unil

Aperitivo con AI - 4 novembre 2020

Partiamo da una domanda

È possibile avere a che fare con un'IA «open-ended»?

- ➤ Open-ended: senza limiti, senza vincoli, aperta al cambiamento, indeterminata
- ➤ Problema della predizione

Previsione e filosofia della scienza

Il problema della predizione riguarda molti campi: probabilità, statistica, teoria dei giochi, causalità, modelli scientifici, approcci simulativi, ecc.

Filosofia della scienza generale: predizione & spiegazione

Q1

Una legge o una teoria scientifica devono permettere di spiegare e predire un fenomeno

Ma...
Spiegazione vs Predizione

Previsione e filosofia della scienza

Limiti della conoscenza scientifica dovuti alla predizione (Rescher 1984, 1999, 2012)

Asimmetria tra spiegazione e predizione da vari punti di vista

- > Logico: deduzione vs induzione; certezza vs probabilità
- ➤ Epistemologico: la spiegazione è una catena causale di fatti, la predizione non può escludere le alternative
- Ontologico: i fatti passati sono chiusi, gli eventi futuri sono aperti

La predizione è limitata rispetto alla spiegazione ed è soggetta a incertezza (razionalità limitata, inaccessibilità cognitiva, problema della rilevanza, ecc.)



Quando usiamo la predizione in IA









TEORIA DELLA

COMPUTAZIONE (PER

PREDIRE SE UN

ALGORITMO SI

FERMERÀ)

TEORIA DEI GIOCHI E
TEORIA DELLA
DECISIONE (PER PREDIRE
IL COMPORTAMENTO
DELL'AGENTE IN UNA
SITUAZIONE
INTERATTIVA O IN UNA
SITUAZIONE DI SCELTA)

APPROCCIO
SIMULATIVO E
METODO SINTETICO
(PER COSTRUIRE E
SFRUTTARE MODELLI E
ARTEFATTI SOFTWARE
E/O HARDWARE)

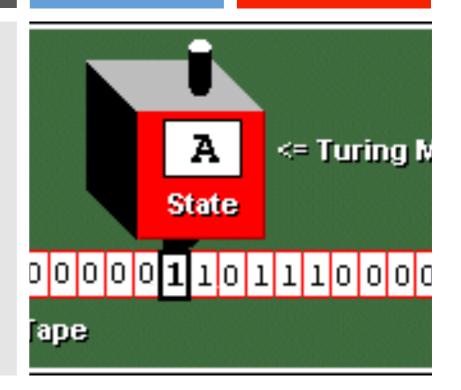
SVILUPPI TECNOLOGICI FUTURI E LORO IMPATTI

Predizione e teoria della computazione

Teoria della computazione: i problemi decidibili sono quelli che sono computabili in un **numero finito di passi**, ma noi non sappiamo quanti passi e quali (→ «opacità algoritmica» generica)

I problemi risolvibili da una Macchina di Turing (Tesi di Church-Turing) necessitano di una quantità finita di memoria \rightarrow l'algoritmo arriva a una conclusione – il programma si ferma – ma più grande è la complessità dell'algoritmo e degli input, più difficile è sapere quando il programma si fermerà e i suoi risultati (problema della trattabilità)

Da un punto di vista empirico, la computazione è **oltre i limiti** della conoscenza umana (→ questione dell'osservatore)



Turing sulla predizione (Turing, 1950)

Risposta all'obiezione di Lady Lovelace: una macchina "has no pretensions to originate anything. It can do whatever we know how to order it to perform."

Turing:

"Machines take me by surprise with great frequency. This is largely because I do not do sufficient calculation to decide what to expect them to do, or rather because, although I do a calculation, I do it in a hurried, slipshod fashion, taking risks. [...] The view that machines cannot give rise to surprises is due, I believe, to a fallacy to which philosophers and mathematicians are particularly subject. This is the assumption that as soon as a fact is presented to a mind all consequences of that fact spring into the mind simultaneously with it. It is a very useful assumption under many circumstances, but one too easily forgets that it is false. A natural consequence of doing so is that one then assumes that there is no virtue in the mere working out of consequences from data and general principles."



Algoritmi, IA e predizione

IA **enfatizza** questo aspetto per due ragioni

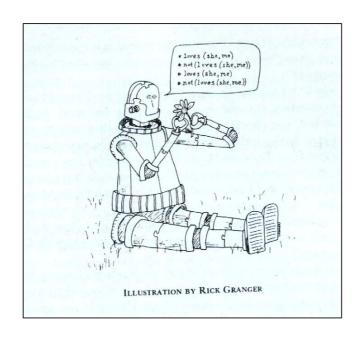
- 1. IA mira a produrre sistemi autonomi
- 2. Gli artefatti dell'IA interagiscono con contesti **open-ended**

Elaborare predizioni è molto difficile a causa della **indeterminatezza** (open-endedness) del **mondo reale**

Sistemi IA che si muovono nel mondo reale (ambienti fisici, sociali, di dati, ecc.) → in larga parte bio-ispirati

La logica dell'IA

Logica come base di partenza per l'IA



Il mondo reale è più complesso della logica

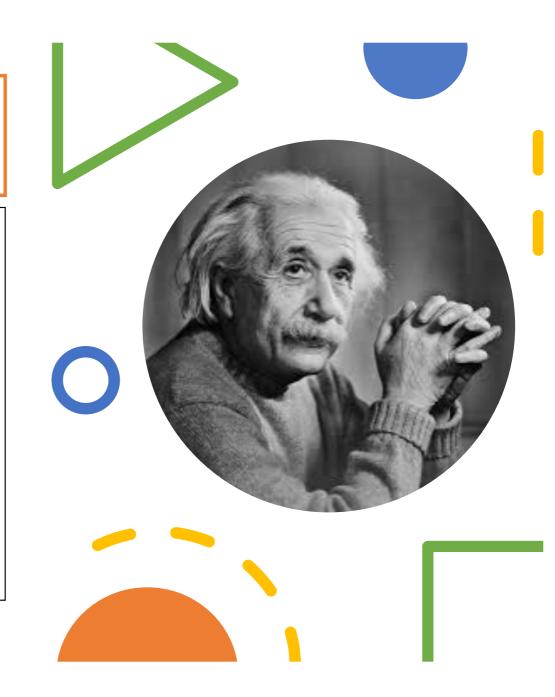
IA e cognizione

Conversazione col cervello di Einstein

[Hofstadter, Dennett, 1981]

Achille: [...] una mente nella sua interazione con il mondo esterno durante un certo intervallo di tempo, si modifica in un modo che non era inizialmente inerente alla sua struttura fisica.

Tartaruga: Osservazione valida, la sua. Una mente, o un cervello, interagisce col mondo ed è quindi soggetto a cambiamenti che non si possono prevedere in base alla sola conoscenza della struttura del cervello.



IA e biologia



«Preludio e... mirmecofuga», in *Gödel, Escher, Bach*, Douglas R. Hofstadter, 1979.

- ☐ Sistemi biologici
- ☐ Sistemi complessi
- ☐ Sistemi autonomi
- ☐ Sistemi che si autoorganizzano
- ☐ Sistemi in grado di adattarsi al mondo esterno (contesto open-ended)

Rapporto fra olismo, emergentismo e riduzionismo: questione dell'osservatore



Intelligenza artificiale bio-ispirata

[Floreano, Mattiussi, 2008]

- ➤ Sistemi evolutivi
- ➤ Sistemi cellulari
- Sistemi neurali
- ➤ Sistemi di sviluppo
- ➤ Sistemi immunitari
- ➤ Sistemi comportamentali
- ➤ Sistemi collettivi (swarm)

Euristiche di ispirazione da differenti sistemi biologici

BIOLOGY, CONTROL, AND ARTIFICIAL INTELLIGENCE

ADAPTATION

IN

NATURAL

AND

ARTIFICIAL

SYSTEMS

JOHN H. HOLLAND

IA, biologia e predizione

Sistemi complessi adattivi (in due sensi)

- 1. Conosciamo le parti ma non il comportamento complessivo
- 2. Conosciamo il punto di partenza, ma non il punto di arrivo, né l'evoluzione

Complessità algoritmica → irriducibilità computazionale

Nessuna scorciatoia!

 Irriducibilità computazionale: per alcuni sistemi fisici è possibile simulare ogni passo dell'evoluzione del comportamento del sistema, ma non è possibile predire un risultato di tale evoluzione senza lasciare che il sistema compia ogni passaggio evolutivo.

Non ci sono scorciatoie per predire il comportamento del sistema!

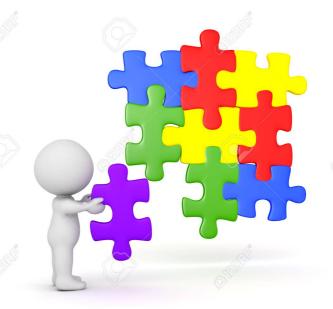
 Tipico dei fenomeni emergenti, come i fenomeni biologici e le loro simulazioni (evolutive e adattive) → esempio degli automi cellulari

IA, biologia e predizione

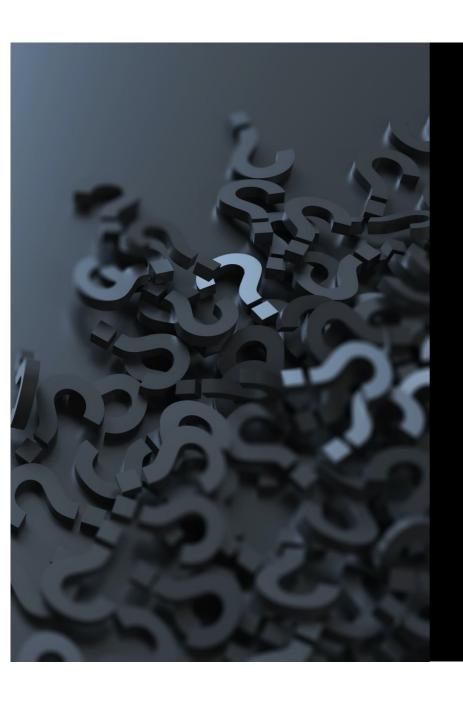
Mettiamo insieme i pezzi...

IA con euristiche bio-ispirate (in senso ampio) è costitutivamente impredicibile.

- a) Il sistema deve evolvere verso la soluzione (non si vuole conoscere in anticipo come)
- b) Il sistema deve adattarsi al contesto openended (non si può conoscere in anticipo come)



- a) e b) sono due casi di libertà nel senso di autonomia.
- → conoscere in anticipo il comportamento del sistema ne mina l'efficacia



E la spiegazione?

Due sensi della spiegazione

1. Possiamo spiegare a ritroso e non in avanti (irriducibilità computazionale) → senso epistemologico di spiegazione

RICOSTRUIRE TUTTI I PASSI COMPIUTI

2. Ma anche potendo spiegare in questo senso, in alcuni casi la spiegazione non è completa, perché i passi seppure conosciuti non sono interpretabili (problema dell'explainable AI)

Predizione e spiegazione

Predizione Spiegazione

- ✓ Impredicibilità costitutiva → Ciò che perdiamo in spiegazione guadagniamo in efficienza
- Risultato emergente è comunque un risultato
- **Ipotesi** per ridurre il problema dell'*Explanable AI* → aumentare la sovrapposizione fra spiegazione e predizione

Conclusioni

- Impredicibilità costitutiva dei sistemi di IA (specialmente bio-ispirati)
- Vantaggi: sistemi autonomi, liberi, adattivi, molto più vicini alla cognizione umana (non monotonicità, contraddizione, individualità)
- Svantaggi: Quanto spiegabili? Quanto controllabili?
- **Paradosso** (dal punto di vista cognitivo): modelli che simulano ma non spiegano

Un'ultima considerazione (provocazione?) sull'intelligenza (artificiale)

Dibattito sulla superintelligenza e/o IA generale

Un sistema adattivo, impredicibile e autonomo potrebbe diventare un'intelligenza con potenzialità, obiettivi, motivazioni e metodi molto lontani dalla comprensione umana?

Esiste il rischio di una IA o di forme di IA (non per forza generali) che non comprendiamo (controlliamo, prediciamo, spieghiamo) che invece possano fare l'inverso (comprendere, predire, controllare **noi**)?

