

Teorema 2 – Teorema Ingenia

AI by Learning este o problemă care cere găsirea unui algoritm A de inteligență artificială care, rulat pe diferite situații date ca input, arată un comportament asemănător cu cel uman. Mai concret, ne dorim să găsim un program care să poată învăța să ia decizii „umane” în orice situație posibilă, dar într-un mod eficient.

Definiție formală:

Se dau:

- un număr întreg K care limitează dimensiunea maximă a algoritmului A
- o metodă de a genera exemple dintr-o distribuție D (care reprezintă posibile comportamente umane într-o varietate de situații).

Să se găsească o descriere L_a (un fel de „manual de funcționare” pentru algoritmul A), astfel încât:

- Dimensiunea acestui manual să fie mai mică sau egală cu K.
- Algoritmul A să performeze cu o probabilitate $\delta(n)$ suficient de mare.

$$\Pr_{s \sim \mathcal{D}_n} [A(s) \in B_s] \geq |B_s|/|B| + \epsilon(n).$$

Aici, δ reprezintă șansa ca algoritmul A să ia o decizie corectă pentru o situație dată și ϵ este o funcție care măsoară cât de „aproape” este comportamentul algoritmului A de cel uman. Acestea sunt funcții arbitrare, dar non-neglijabile. O funcție f este considerată non-neglijabilă dacă există un d astfel încât, pentru valori suficient de mari ale lui n , $f(n) \geq 1/n^d$.

Teorema Ingenia spune că găsirea unui astfel de algoritm care să poată rezolva orice sarcină imitând comportamentul uman nu este posibilă în timp polinomial. Pe scurt, “AI by Learning” este intractable.

Pentru a demonstra teorema ne folosim de rezultatul anterior, și anume că problema Perfect vs Chance este intractable:

1. Presupunem că există un mecanism de învățare AI, M, care poate învăța eficient orice sarcină (adică AI-by-Learning este tractabil). Folosind acest mecanism, putem rezolva problema Perfect-vs-Chance într-un mod eficient. În această problemă, ne este dată o distribuție de date D, iar mecanismul M învață un algoritm A care încearcă să prezică ieșirile y pe baza intrărilor x . Apoi, folosim noi exemple pentru a testa cât de bine funcționează A. Dacă A prezice corect aproape întotdeauna,

- înseamnă că există un program perfect (Cazul 1), iar dacă A nu face mai mult decât ghicitul aleatoriu, înseamnă că nu există un program bun (Cazul 2). După ce rulăm acest mecanism de mai multe ori și testăm diferiți algoritmi A, alegem pe cel care obține cele mai bune rezultate. Dacă acest algoritm performează semnificativ mai bine decât ghicitul aleatoriu, declarăm că suntem în Cazul 1, iar dacă nu, declarăm Cazul 2. Astfel, am rezolvat problema Perfect-vs-Chance într-un timp polinomial.
2. Știm din rezultatul anterior că Perfect vs Chance este intractable. Astfel, știind că $p \rightarrow q \Leftrightarrow !q \rightarrow !p$, rezultă că și AI-by-Learning este intractable. Contradicție.

Din 1. și 2. Rezultă că AI-by-Learning este intractable.

Teorema Ingenia ne arată limitele fundamentale ale inteligenței artificiale bazate pe învățare. Deși putem dezvolta algoritmi care să performeze bine în sarcini specifice, ideea unui algoritm universal, care să învețe orice sarcină umană în mod eficient, este imposibilă din punct de vedere teoretic. Din acest motiv o abordare mai utilă ar fi întoarcerea la ideea mai veche de a folosi inteligența artificială pe post de unealtă în domeniul științei cognitive.