3. Arquitectures de la IA

I quina de les arquitectures hauria d'usar un agent? La resposta és que totes elles. Calen respostes reflexes a les situacions en què el temps és essencial, mentre que la deliberació basada en el coneixement permet a l'agent planificar amb antelació. L'aprenentatge convé quan es disposa d'una gran quantitat de dades, i és necessari quan l'entorn canvia, o quan els dissenyadors humans tenen un coneixement insuficient del domini.

La IA ha tengut de fa temps una escletxa entre els sistemes simbòlics (basats en la lògica i en la inferència probabilística) i els sistemes connexionistes (basats en la minimització d'una pèrdua sobre un gran nombre de paràmetres sense interpretació). Un repte continu a la IA és ajuntar aquests dos enfocaments, per capturar la millor part de tots dos. Els sistemes simbòlics permeten concatenar llargues cadenes de raonament i aprofitar l'avantatge de la potència expressiva de les representacions estructurades, mentre que els sistemes connexionistes poden reconèixer patrons fins i tot en presència de dades sorolloses. Una línia de recerca intenta de combinar programació probabilística amb aprenentatge profund.

Els agents també necessiten maneres de controlar les seves pròpies deliberacions. Han de ser capaços d'usar bé el temps disponible, i deixar de deliberar quan cal actuar. Per exemple, un agent que mena un taxi i troba un accident ha de decidir en una fracció de segon si ha de frenar o virar. També ha de dedicar aquesta fracció de segon a les qüestions més importants, com ara si les línies esquerra i dreta de la via són clares o si s'acosta un gran camió per darrere, abans de preocupar-se d'on agafar el proper passatger. Aquestes qüestions entren dins l'àrea de la IA en temps real. A mesura que els sistemes entren en dominis més complexos, tots els problemes esdevenen de temps real, perquè l'agent mai no tendrà prou temps per resoldre exactament el problema.

Hi ha una necessitat de mètodes generals de control de la deliberació, més que no de receptes específiques sobre què pensar en cada situació. Una primera idea útil són els algorismes anytime, que es poden interrompre en qualsevol moment i donen una solució aproximada, i ofereixen una solució de més qualitat com més temps de càlcul usen.

Una segona tècnica per controlar la deliberació és el **meta-raonament** basat en la **teoria de la decisió**. Aquest mètode aplica la teoria del valor de la informació a la selecció dels càlculs individuals. El valor d'un càlcul depèn del seu cost (en termes de retard a l'acció) i els seus beneficis (en termes de la qualitat millorada de la decisió). La tècnica de meta-raonament es pot usar per dissenyar algorismes de cerca més bons i per garantir que els algorismes tenen la propietat anytime.

El meta-raonament és un exemple d'arquitectura reflexiva, que permet la deliberació sobre les entitats i accions computacionals que s'esdevenen dins l'arquitectura mateixa.

A continuació desenvolupam els tres apartats següents.

1. IA general
2. Enginyeria de la IA
3. El futur

3.1. IA general

Molta part del progrés de la IA el segle XXI ha estat guiat per la competició en tasques concretes, com el DARPA Grand Challenge per a vehicles autònoms, la competició de reconeixement d'objectes ImageNet, o els jocs de Go, escacs, pòquer o Jeopardy! contra un campió del món. Per a cada tasca individual, es construeix un sistema diferent d'IA, habitualment un model d'aprenentatge automàtic entrenat des de zero amb dades recollides específicament per a aquella tasca. Però un agent veritablement intel·ligent hauria de ser capaç de fer més d'una cosa.

Alan Turing va proposar la seva llista i l'autor de ciència ficció Robert Herlein la seva (1973):

> Un humà hauria de ser capaç de canviar un bolquer, planificar una invasió, matar un porc, agafar un vaixell, dissenyar un edifici, escriure un sonet, quadrar comptes, aixecar una paret, posar a lloc un os trencat, consolar un moribund, obeir ordres, donar ordres, cooperar, actuar tot sol, resoldre equacions, analitzar un problema nou, tirar els fems, programar una computadora, cuinar, lluitar, morir dignament. L'especialització és per als insectes.

Fins ara, cap sistema no és capaç de fer totes aquestes coses. Alguns proponents de la IA general o de nivell humà (HLAI, Human Level Artificial Intelligence) insisteixen que la feina continuada en tasques específiques (o components específics) no bastarà per dominar una àmplia varietat de tasques, sinó que caldrà un enfocament fonamentalment nou. Segons Russell i Norvig, seran necessaris nous avenços, però el camp de la IA ha fet un equilibri raonable entre exploració i explotació, reunint un portafoli de components, millorant tasques particulars, alhora que explorant idees prometedores i de vegades noves i llunyanes.

Hem vist com la feina en els components pot generar noves idees; per exemple, les xarxes antagonistes generatives (GAN) i els transformers han obert àrees noves de recerca. També hem vist passes cap a una diversitat de comportament. Per exemple, els sistemes de traducció automàtica de la dècada dels 90 es construïen per a una parella de llengües cada vegada, mentre que ara poden traduir entre qualsevol parell entre més de cent llengües.

3.2. Enginyeria de la IA

El camp de la programació de computadores començà amb un grapat de pioners extraordinaris. Però no va assolir l'estatus d'una indústria important fins que es va desenvolupar una pràctica professional d'enginyeria del programari, amb una gran col·lecció d'eines disponibles per a tothom, i un ecosistema dinàmic de professors, estudiants, professionals, emprenedors, inversors i clients.

La indústria de la IA encara no ha assolit aquest nivell de maduresa. Hi ha una sèrie d'eines i *frameworks*, com TensorFlow (amb Keras), PyTorch, CAFFE, scikit-learn o SciPy. però moltes dels enfocaments prometedors, com les GAN o l'aprenentatge de reforç profund, han demostrat esser difícils de fer-hi feina. Cal experiència i un cert grau d'experimentació per aconseguir-los traslladar a dominis nous. Encara no hi ha prou experts per fer-ho en tots els dominis on calen, i encara no hi ha les eines i l'ecosistema perquè els professionals més poc experts se'n surtin.

Jeff Dean, cap de Google AI del 2018 al 2023, preveu un futur on voldrem que l'aprenentatge automàtic realitzi milions de tasques. No serà factible desenvolupar-les cadascuna de zero, de forma que la seva suggerència és començar amb un únic sistema gegant i, per a cada nova tasca, extraure'n les parts rellevants per a la tasca. Ja s'han vist algunes passes en aquesta direcció, com els models de llenguatge *transformer* (BERT, GPT) amb bilions de paràmetres.

Amb les diverses iteracions de GPT (al moment d'escriure aquests apunts, 2024, GPT-4) aquesta visió es concreta en el *fine-tuning*, que adapta un model de llenguatge massiu a la tasca concreta que s'hi vulgui realitzar.

3.3. El futur

Russell i Norvig acabaven la quarta edició del seu llibre Artificial Intelligence, l'any 2022, amb la reflexió que donam traduïda a continuació.

Cap a on anirà el futur? Els autors de ciència ficció sembla que prefereixen els futurs distòpis respecte dels utòpics, probablement perquè això genera trames més interessants. Fins ara, la IA sembla que està al nivell d'altres tecnologies poderoses i revolucionàries com la impremta, l'aigua corrent (*plumbing*), el transport aeri i la telefonia. Totes aquestes tecnologies han tengut impactes positius, però també tenen efectes col·laterals imprevistos que tenen un impacte desproporcionat sobre les classes desfavorides. Faríem bé d'invertir en minimitzar els impactes negatius.

La IA també és diferent de les tecnologies revolucionàries precedents. Millorar la impremta, l'aigua corrent, el transport aeri o la telefonia fins als seus límits no produeix res que amenaci la supremacia humana sobre el món. Millorar la IA fins al seu límit lògic sí que ho fa.

Com a conclusió, la IA ha fet un gran progrés en la seva breu història, però la frase final de l'assaig d'Alan Turing el 1950, Computing Machinery and Intelligence, encara és vàlida avui:

We can only see a short distance ahead, but we can see that much remains to be done.

Només podem veure-hi una curta distància cap endavant, però podem veure que hi manca molt a fer.