
Què és la intel·ligència artificial

PID_00267994

Vicenç Torra i Reventós

Temps mínim de dedicació recomanat: 2 hores



Vicenç Torra i Reventós

La revisió d'aquest recurs d'aprenentatge UOC ha estat coordinada pel professor: Carles Ventura Royo (2019)

Tercera edició: setembre 2019
© Vicenç Torra i Reventós
Tots els drets reservats
© d'aquesta edició, FUOC, 2019
Av. Tibidabo, 39-43, 08035 Barcelona
Realització editorial: FUOC

Cap part d'aquesta publicació, incloent-hi el disseny general i la coberta, no pot ser copiada, reproduïda, emmagatzemada o transmesa de cap manera ni per cap mitjà, tant si és elèctric com químic, mecànic, òptic, de gravació, de fotocòpia o per altres mètodes, sense l'autorització prèvia per escrit dels titulars del copyright.

Índex

Introducció	5
Objectius	6
1. Breu repàs històric de la intel·ligència artificial	7
2. Què és la intel·ligència artificial: definicions i punts de vista	11
2.1. La hipòtesi del sistema de símbols físics	13
2.2. Sistemes basats en models biològics	15
2.3. Sistemes emergents	16
2.4. Sistemes basats en les teories de l'acció situada i la intel·ligència artificial corpòria	17
2.5. Mètodes i tècniques de la intel·ligència artificial	19
3. Algunes aplicacions	21
4. Algunes característiques dels sistemes intel·ligents	25
Activitats	27
Bibliografia	28

Introducció

Aquest mòdul presenta una visió general de la intel·ligència artificial i introdueix l'assignatura. Per tal de donar un aspecte global de l'àrea i de les aplicacions existents, el mòdul s'ha dividit en quatre apartats.

Comencem amb un apartat dedicat a fer un breu repàs de la història de l'àrea a partir d'alguns dels esdeveniments més importants.

A continuació, es dedica un apartat a veure què és la intel·ligència artificial. Atès que l'opinió dels investigadors no coincideix a l'hora de definir-la ni de dir quins en són els objectius principals, presentem diversos punts de vista. La descripció ens permet d'enquadrar l'assignatura dins el cos general de la intel·ligència artificial.

El tercer apartat està dedicat a descriure algunes de les aplicacions existents, com ara demostradors de teories o sistemes experts.

El mòdul acaba amb un apartat que inclou un repàs de les característiques que comparteixen els sistemes que fan servir tècniques d'intel·ligència artificial.

Objectius

Aquest mòdul didàctic vol assolir els objectius següents:

- 1.** Conèixer els diferents punts de vista que hi ha sobre la intel·ligència artificial.
- 2.** Saber emmarcar el temari de l'assignatura dintre de l'àrea.
- 3.** Tenir una idea general de la història de la intel·ligència artificial.
- 4.** Conèixer alguns exemples pràctics en què s'apliquen les tècniques d'intel·ligència artificial a problemes reals.

1. Breu repàs històric de la intel·ligència artificial

El terme *intel·ligència artificial* (IA) va ser adoptat l'estiu del 1956 a Dartmouth en una estada que va reunir investigadors interessats en els temes de la intel·ligència, les xarxes neurals i la teoria d'autòmats. Aquesta estada va ser organitzada per J. McCarthy (que era a Dartmouth), C. Shannon, M. Minsky i N. Rochester i hi van assistir T. More (Princeton), A. Samuel (IBM), R. Solomonoff i O. Selfridge (MIT) i A. Newell, H. Simon (Carnegie Tech, actualment Carnegie Mellon University). La reunió va constituir el que després seria el nucli central de la intel·ligència artificial.

Notació

Abreugem el terme *intel·ligència artificial* amb la sigla IA.

La trobada va ser la conseqüència dels primers treballs que s'estaven fent en el camp. Per exemple, Newell i Simon hi van presentar el seu programa de raonament automàtic Logic Theorist (que va causar molta sensació). De fet, actualment es considera que la intel·ligència artificial neix amb un treball desenvolupat per W. McCulloch i M. Pitts del 1943 en què es proposava un model de neurones artificials basat en tres fonts: coneixements sobre la fisiologia i el funcionament de les neurones del cervell; la lògica proposicional de Russell i Whitehead a més de la teoria de la computació de Turing. D'uns anys més tard fou el primer ordinador neural (l'SNARC). Fou construït cap allà l'any 1951 per dos estudiants de Princeton: M. Minsky, i D. Edmons. Més o menys de la mateixa època foren els primers programes per a jugar als escacs de Shannon i Turing.

Tot i que aquests són els treballs amb els quals s'inicia la intel·ligència artificial, hi ha molts treballs anteriors que han influït en l'evolució d'aquesta àrea. Alguns d'aquests treballs han aparegut en els camps de la filosofia (un dels primers intents de formalitzar el raonament són els sil·logismes del filòsof grec Aristòtil), les matemàtiques (la teoria de la decisió i la teoria de la probabilitat s'apliquen en molts dels sistemes actuals) o la psicologia (que amb la intel·ligència artificial han definit el camp de la ciència cognitiva).

En els anys següents a la reunió de Dartmouth es van aconseguir molts progressos. Es van construir programes que resolien problemes molt diversos. Per exemple, estudiants de Minsky implementaren al final de la dècada dels seixanta els programes *Analogy*, que resolia problemes geomètrics com els que apareixen en els tests d'intel·ligència, i *Student*, que resolia problemes d'àlgebra redactats en anglès. Newell i Simon desenvoluparen el *General Problem Solver* (GPS), que imitava el raonament humà. Samuel va escriure programes per a jugar als *checkers* (joc semblant a les dames) que aprenien a jugar més bé que l'autor. McCarthy, que havia marxat al MIT, va implementar l'any

1958 el llenguatge Lisp i va escriure aquell mateix any un treball titulat *Programs with Common Sense* en què es descriu un programa hipotètic que constitueix el primer sistema complet d'intel·ligència artificial.

Aquesta sèrie d'èxits queda truncada a mitjan dècada dels seixanta i les previsions, massa optimistes, d'aquells anys es frustraren. Els sistemes implementats fins a aquell moment funcionaven en dominis limitats (coneguts com a micromons¹). La transformació per a adaptar-se a entorns reals no va ser (no ha estat) trivial com s'esperava. Això va ser degut, fonamentalment, segons Russell i Norvig, als tres factors següents:

1) Alguns dels sistemes no disposaven de coneixement de l'entorn de l'aplicació o aquest era molt petit i únicament consistia en algunes manipulacions sintàctiques simples.

ELIZA: un sistema de manipulacions sintàctiques simples

El programa de diàlegs ELIZA de Weizenbaum volia simular un psicòleg que conversava amb el pacient. El seu funcionament es basava en el reconeixement d'estructures en les frases del pacient i sobre la base d'uns patrons construïa unes noves frases. Tenia regles com les dues que donem a continuació (on per a cada regla hi ha el patró i una possible resposta o més):

```
(((* x) recordo (* y))
 (penses sovint en (* y) ?)
 (((* x) he somniat (* y))
 (segur-- (* y) ?)
 (i ara per què em parles de somnis ?)
 (has somniat amb (* y) abans ?))
```

Amb aquestes regles, quan algú escriu "ara encara recordo el viatge a París", ELIZA respon "penses sovint en el viatge a París?" perquè reconeix l'estructura "recordo" i associa el que segueix a aquest mot a la variable *y* i després construeix una nova frase amb el contingut d'aquesta *y*. El mateix farà amb la segona regla. En aquest cas, hi ha tres possibles respostes.

2) Molts dels problemes que s'intentaven resoldre eren problemes intractables (NP-complexos). Mentre la quantitat de coneixement era petita es podien resoldre, però quan aquesta era més gran els problemes resultaven irresolubles.

Exemple de problema intractable

Un problema que considera permutacions d'objectes en passar de 8 a 10 objectes triga unes 100 vegades més. Si passem a 20 objectes són 6×10^{13} vegades més!

3) Algunes de les estructures bàsiques que s'utilitzaven per a generar una conducta intel·ligent tenien grans limitacions.

Limitacions dels perceptrons

Els perceptrons (un tipus de xarxa neural) desenvolupats per Minsky i Papert no podien aprendre la funció *or* exclusiva (*xor*). Això és, no hi ha cap perceptró (cap assignació de pesos a les connexions) amb dues unitats d'entrada i una de sortida que es comporti com la funció *or* exclusiva. És a dir, que quan les dues entrades són zero, el resultat sigui zero; que si les dues són 1, també doni 0; i que quan una de les dues és 1, el resultat sigui 1.

⁽¹⁾En anglès, *microworlds*.

Lectura complementària

Vegeu les principals dificultats d'adaptació dels sistemes intel·ligents a entorns reals al primer capítol de l'obra següent:

S. Russell; P. Norvig (2003). *Artificial Intelligence: A Modern Approach* (2^a ed.). Englewood Cliffs: Prentice Hall.

Algorismes NP

De manera informal definim un algorisme NP com un algorisme que necessita un temps que creix exponencialment amb la mida de les dades d'entrada.

Entrada	Entrada	Funció xor
0	0	0
0	1	1
1	0	1
1	1	0

Per tant, els perceptrons no servien per a qualsevol tipus de problema. En particular, no serveixen per als problemes de classificació que no són linealment separables.

Fins aquell moment la resolució de problemes es basava en un mecanisme de cerca de propòsit general que intentava encadenar passos elementals de raonament per a trobar la solució completa. El cost d'aquest mecanisme era molt elevat; per a reduir-lo es van desenvolupar els primers algorismes de cerca. Per exemple, Dijkstra l'any 1959 va dissenyar el mètode de cost uniforme, Newell i Ernst l'any 1965 van desenvolupar el concepte de cerca heurística i Hart, Nilsson i Raphael, l'algorisme A*. En aquesta mateixa època, en relació amb els programes de jocs, es definia la cerca alfa-beta. De fet, va ser ideada per McCarthy l'any 1956 i utilitzada més tard (l'any 1958) per Newell.

La importància de considerar el coneixement del domini de l'aplicació i la seva estructuració per tal que el seu accés fos eficient i fes possible la construcció de sistemes per a entorns reals va portar a l'estudi dels mètodes de representació del coneixement. Entre d'altres, es van definir les xarxes semàntiques (definides al principi dels anys seixanta per Quillian) i els marcs (definites per Minsky l'any 1975). En aquest mateix període es van començar a fer servir alguns tipus de lògica per a representar el coneixement.

De manera paral·lela, durant aquests anys continuen els treballs, en la línia del sistema per a jugar als *checkers* desenvolupat per Samuel, orientats a incorporar-hi algun tipus d'aprenentatge. E.B. Hunt, J. Marin i P.T. Stone construeixen l'any 1966 els arbres de decisió (per a classificació), ja ideats per Shannon l'any 1949, i Quinlan (1979) presenta el mètode ID3 per a construir aquests arbres. D'altra banda, P. Winston (1970) desenvolupa el seu programa per a aprendre descripcions d'objectes complexos i T. Mitchell (1977) desenvolupa l'anomenat *espai de versions*. Més tard, a mitjan dècada dels vuitanta, el redescobrimient del mètode d'aprenentatge per a xarxes neuronals anomenat *retropropagació* ("backpropagation") porta al ressorgiment d'aquest camp.

La construcció d'aplicacions per a entorns reals ha obligat a tractar aspectes com la incertesa o la imprecisió (necessaris també per a resoldre els jocs d'atzar). S'han aplicat mètodes probabilístics (teoria de la probabilitat o xarxes probabilístiques) i desenvolupat altres formalismes com els conjunts difusos (definites per L. Zadeh l'any 1965) o la teoria de Dempster-Shafer (original d'A. Dempster l'any 1968 amb aportacions de G. Shafer de l'any 1976).

Vegeu també

Vegeu els algorismes de cost uniforme i A* en els subapartats 4.1 i 4.3 del mòdul "Resolució de problemes i cerca" d'aquesta assignatura.

Vegeu també

L'algorisme de retropropagació i els models basats en xarxes neuronals s'introdueixen en el mòdul 5 "Introducció a l'aprenentatge computacional" d'aquesta assignatura i es veuran amb més detall a l'assignatura d'*Aprenentatge computacional*.

Vegeu també

Alguns dels aspectes relacionats amb la incertesa i la imprecisió es descriuen en el mòdul "Incertesa i raonament aproximat" d'aquesta assignatura.

Amb aquests desenvolupaments es van construir a partir dels anys vuitanta els primers sistemes comercials (principalment sistemes experts).

Avui en dia els problemes de recerca en el camp de la intel·ligència artificial van orientats cap a la creació de sistemes cooperatius basats en agents, incloent-hi sistemes per a la recuperació i gestió de dades, visió per computador i processament de llenguatge natural, entre d'altres.

2. Què és la intel·ligència artificial: definicions i punts de vista

A diferència del que passa en altres àrees, en la intel·ligència artificial no hi ha una única definició, sinó que n'hi ha moltes corresponents a punts de vista diferents. En aquest apartat comencem amb un petit repàs d'algunes de les definicions existents. Després es veuran les hipòtesis sobre les quals s'han desenvolupat els programes.

Construcció de programes

La intel·ligència artificial, com a camp de la informàtica que és, està dedicada a la construcció de programes.

Utilitzant una frase de Simon podem dir que:

“The moment of the truth is a running program”.

H. A. Simon. *Artificial intelligence* (1995, pàg. 96).

Conducta enfront de raonament

De tota manera, els objectius que persegueixen aquests programes ens portaran a diferents definicions. En el seu llibre d'intel·ligència artificial, Russell i Norvig classifiquen els objectius d'acord amb dues dimensions:

Lectura complementària

Trobareu la classificació dels objectius de la IA segons Russell i Norvig en el primer capítol de l'obra següent:

S. Russell; P. Norvig (2003). *Artificial Intelligence: A Modern Approach* (2^a ed.). Englewood Cliffs: Prentice Hall.

1) D'una banda, l'objectiu dels programes pot ser obtenir una conducta determinada (ens interessa el resultat que s'obté – el comportament) o bé aconseguir una manera de raonar determinada (més que el resultat el que ens interessa és que el raonament d'una matèria es faci d'una manera determinada). Aquest parell conducta/raonament constitueix la primera dimensió.

2) De l'altra, s'exigirà que els programes siguin correctes. Aquesta correcció es pot mesurar comparant el rendiment amb el de les persones o bé comparant-lo amb un concepte ideal d'intel·ligència. Aquest ideal l'anomenem *racionalitat*. La forma de mesurar la correcció defineix la segona dimensió.

D'acord amb aquestes dues dimensions hi ha quatre visions possibles per als objectius i, per tant, quatre definicions possibles per al terme *intel·ligència artificial*. Aquestes alternatives han estat considerades al llarg de la història del camp i són aquestes:

a) Actuar com les persones: el test de Turing

Aquest enfocament està basat en la proposta d'A. Turing que defineix la intel·ligència de manera operacional: una conducta és intel·ligent quan té prou nivell per a confondre un interlocutor humà. La descripció del test apareix en un article de l'any 1950 de la revista *Mind* (núm. 59, pàg. 433-60).

b) Raonar com les persones: el model cognoscitiu

En aquest enfocament no interessa únicament el resultat, sinó també com s'obté. Es vol que el procés de raonament sigui semblant al de les persones. El primer sistema amb aquests objectius fou el General Problem Solver de Newell i Simon de l'any 1961. El camp de la ciència cognitiva utilitza aquest punt de vista.

c) Raonar racionalment: les lleis del pensament

Aquest punt de vista es basa a plantejar què significa pensar correctament o, el que és equivalent, saber quan es pot deduir de manera lògica un fet nou a partir del coneixement de què es disposa. Els sistemes que utilitzen algun tipus de lògica es construeixen sota aquest punt de vista.

d) Actuar racionalment: l'agent racional

Actualment el terme *agent* està molt difós. Sobre què és un agent, l'Artificial Intelligence FAQ diu:

A very misused term. Today, an agent seems to mean a stand-alone piece of AI-ish software that scours across the internet doing something "intelligent". Russell and Norvig define it as "anything that can be viewed as perceiving its environment through sensors and acting upon that environment through effectors." Several papers I've read treat it as 'any program that operates on behalf of a human,' similar to its use in the phrase 'travel agent'. Marvin Minsky has yet another definition in the book "Society of Mind." Minsky's hypothesis is that a large number of seemingly-mindless agents can work together in a society to create an intelligent society of mind. Minsky theorizes that not only will this be the basis of computer intelligence, but it is also an explanation of how human intelligence works.

No és important que les conclusions a què arribi un sistema s'obtinguin a partir de deduccions lògiques (inferències correctes) del coneixement, sinó que s'aconsegueixin els objectius (a partir de les pròpies creences – d'allò que es creu que és cert). A més, l'avaluació d'aquests objectius (l'aspecte de la correcció) es fa utilitzant un concepte general de racionalitat que es pot definir de manera explícita.

Lectura complementària

Es pot trobar una còpia del treball de Turing esmentat a l'adreça següent:

<http://www.abelard.org/turpap/turpap.htm>

Exemple d'actuació racional

En un programa de jocs, la racionalitat seria que el programa de jocs guanyés sempre amb independència de si les persones guanyen o no.

Quan comparem aquest darrer punt de vista amb el de les lleis del pensament, hem de subratllar que ara la intel·ligència artificial no està limitada a la inferència lògica. Podem fer servir altres mecanismes si són adequats per a aconseguir els resultats.

Ciència i enginyeria

S'ha comentat anteriorment que la intel·ligència artificial es dedica a la construcció de programes. Des d'aquest punt de vista, i més quan tenim en compte que no s'escriuen sols programes sinó que també es dissenyen mecanismes com robots, el camp s'ha de veure com una enginyeria. Tanmateix, no és solament enginyeria, sinó també una ciència. Això és degut al fet que a més de desenvolupar sistemes nous també s'estudia i s'obté nou coneixement sobre els sistemes intel·ligents.

A més, com subratlla Simon, és una ciència experimental atès que s'experimenta quan dissenyem un sistema i observem com canvia en la conducta si canvia el disseny. També s'experimenta quan es fa servir la intel·ligència artificial per a estudiar la intel·ligència natural.

La dualitat del camp explica l'existència en la literatura tant de treballs orientats cap a les qüestions pràctiques com els orientats als aspectes formals. Aquests darrers són necessaris per a construir i explicar el comportament dels primers.

Vegem ara les hipòtesis que són el fonament de la majoria de sistemes en intel·ligència artificial.

2.1. La hipòtesi del sistema de símbols físics

En el centre de la majoria dels treballs en intel·ligència artificial hi ha l'anomenada *hipòtesi del sistema de símbols físics* d'A. Newell i H.A. Simon:

The physical symbol system hypothesis:

A physical symbol system has the necessary and sufficient means for general intelligent action.

A. Newell; H. A. Simon. "Computer Science as Empirical Inquiry: Symbols and Search" (1976).

Un **sistema de símbols físics** disposa d'un conjunt d'entitats, els símbols, que són patrons físics. A més, hi ha un conjunt de processos (operacions) que operen sobre expressions per a produir expressions noves.

Lectura complementària

Trobareu el parer de Simon sobre la IA com a enginyeria en l'obra següent:

H. A. Simon (1995). "Artificial Intelligence: an Empirical Science". *Artificial Intelligence* (núm. 77, pàg. 95-127).

A causa que els sistemes tenen recursos limitats, Newell i Simon consideren que s'han de fer cerques heurístiques per a resoldre els problemes. Aquesta hipòtesi és l'anomenada *hipòtesi de la cerca heurística*:

Heuristic Search Hypothesis:

The solutions to problems are represented as symbol structures. A physical symbol system exercises its intelligence in problem solving by search – that is, by generating and progressively modifying symbol structures until it produces a solution structure.

A. Newell; H. A. Simon. "Computer Science as Empirical Inquiry: Symbols and Search" (1976).

Considerant les dues hipòtesis tenim que l'activitat intel·ligent s'aconsegueix mitjançant l'ús de:

- Símbols: representen els elements significatius d'un problema.
- Operacions: que operen sobre els símbols i permeten generar les solucions.
- Cerca: per a seleccionar una solució entre les solucions possibles que es poden generar a partir de les operacions.

Les dues hipòtesis anteriors justifiquen els esforços per a construir màquines intel·ligents i fan explícites les suposicions que formen la base de la intel·ligència artificial. En particular, permeten distingir entre els patrons de símbols i el mitjà que es fa servir per a implementar-los. Si la intel·ligència només s'obté de l'estructura d'un sistema simbòlic, llavors qualsevol mitjà que implementi de manera adequada aquests patrons i els processos que hi operen podrà aconseguir la intel·ligència amb independència de si el sistema està format de neurones o de circuits lògics.

La hipòtesi del sistema de símbols físics explica els dos elements més importants de la investigació i el desenvolupament d'aplicacions en IA:

- 1) Definir les estructures simbòliques i les operacions necessàries per a resoldre problemes.
- 2) Desenvolupar estratègies per a cercar, de manera correcta i eficient, les solucions potencials generades per aquestes estructures i operacions.

Aquests són els aspectes interrelacionats de cerca i representació del coneixement. Aquests dos elements constitueixen el nucli de l'àrea.

L'opinió de Newell i Simon és, tanmateix, discutida per molts autors dins la intel·ligència artificial. Alguns han construït sistemes que es basen en hipòtesis diferents. Podem subratllar tres alternatives diferents basades en tres models diferents de la intel·ligència. A continuació es presenten les tres alternatives

Vegeu també

El tema de la cerca es veu en el mòdul "Resolució de problemes i cerca" i la representació es veu en el mòdul "Sistemes basats en el coneixement" d'aquesta assignatura.

de manera independent encara que, de fet, hi ha fortes relacions entre si: els sistemes basats en models biològics, els sistemes emergents i els sistemes basats en les teories de l'acció situada i la intel·ligència artificial corpòria.

2.2. Sistemes basats en models biològics

Una alternativa als sistemes de símbols físics són les xarxes neurals, que han estat dissenyades sota la inspiració de models biològics.

Una **xarxa neural** està formada per “neurons” (unitats de processament) i unes connexions entre si que tenen associades uns pesos.

Aquestes xarxes representen de manera implícita el món mitjançant els pesos associats a cada connexió. Per tant, no es fan servir símbols. A més, a causa que el coneixement està distribuït per tota l'estructura de la xarxa és molt difícil, si no impossible, aïllar els conceptes individuals en regions específiques de la xarxa. De fet, una porció de la xarxa es pot fer servir per a representar diversos conceptes.

Les xarxes desplacen l'èmfasi de la intel·ligència artificial fora dels problemes de la representació simbòlica i de les estratègies d'inferència sòlida a problemes d'aprenentatge i adaptació. Les xarxes neurals són mecanismes que s'adapten al món (adopten la seva forma per la interacció amb el món mitjançant tècniques d'aprenentatge que modifiquen els pesos de les connexions), i no necessiten que el món sigui refós en un model simbòlic explícit.

Per tot això, les xarxes neurals es poden veure (com a mínim) com un contra-exemple de la condició de necessitat de la hipòtesi dels sistemes de símbols físics.

Encara que la majoria d'autors consideren que les xarxes vulneren la hipòtesi de Newell i Simon, aquest darrer ha comentat darrerament que cal considerar la condició de símbol àmpliament de manera que els sistemes connexionistes, i també els robots d'R. Brooks es poden incloure en els sistemes de símbols físics:

My work fits within the framework described above in terms of situatedness, embodiment, intelligence and emergence. In particular I have advocated situatedness, embodiment, and highly reactive architectures with no reasoning systems, no manipulable representations, no symbols, and totally decentralized computation. This different model of computation has lead to radically different models of thought.

R. A. Brooks. “Intelligence without Reason” (1991, pàg. 585).

D'aquesta manera la hipòtesi de sistemes de símbols físics es continua considerant vàlida. Sota aquesta consideració, les xarxes neurals es poden veure com un mecanisme més de representació del coneixement.

Vegeu també

Els models basats en xarxes neurals s'introdueixen en el mòdul 5 “Introducció a l'aprenentatge computacional” d'aquesta assignatura i es veuen amb més detall en l'assignatura d'*Aprenentatge computacional*.

There is some dispute today about the Physical Symbol System Hypothesis, hinging on the definition of the term "symbol". If we define "symbol" narrowly, so that the basic components in connectionist systems or robots of the sort advocated by Brooks are not regarded as symbols, then the hypothesis is clearly wrong, for systems of these sorts exhibit intelligence. If we define symbols (as I have, above) as patterns that denote, then connectionist systems and Brooks' robots qualify as physical symbol systems.

H. A. Simon. "Artificial Intelligence: an Empirical Science" (1995, pàg. 104-105).

Com es veurà, en el desenvolupament de l'assignatura es farà servir el punt de vista de Newell i Simon. Això és, es presentaran els elements bàsics de la intel·ligència artificial centrats en els temes de cerca i representació del coneixement. A més, també seguint aquests autors, es veuran les xarxes neurals com una forma de representació del coneixement.

2.3. Sistemes emergents

Una altra alternativa a la hipòtesi de Newell i Simon és la plantejada per les teories emergents.

Es parteix del fet que la solució a un problema no s'obté d'un únic sistema individual sinó que emergeix de les activitats d'uns agents independents (i relativament simples encara que normalment especialistes).

Les tecnologies d'agents, que actualment estan en auge i influeixen en el disseny de sistemes de computació distribuïda (en especial d'Internet), ofereixen models de cooperació social, de manera que es poden construir sistemes complexos a partir de la cooperació d'agents individuals.

Aquestes idees apareixen també en les ciències cognitives, on s'afirma que la ment s'organitza en un conjunt d'unitats funcionals especialitzades.

Aplicacions

L'emergència s'aplica a diversos nivells. Està en l'àmbit d'individus on el comportament d'un individu s'obté a partir de la interacció de mòduls independents, o bé en l'àmbit de poblacions (conjunt d'individus). Aquest darrer cas és el que trobem en els robots dissenyats per l'equip d'R. Brooks que hem esmentat abans. S'han dissenyat i construït robots simples que interaccionen com a agents autònoms solucionant problemes en un laboratori. No hi ha en aquests un algorisme central, sinó que la cooperació emergeix com a resultat de les interaccions distribuïdes i autònomes dels individus. Aplicacions semblants es desenvolupen a Internet. Aquí tenim que les unitats independents són els agents. Cadascun amb les seves pròpies capacitats. És el conjunt d'agents el que defineix la conducta d'un sistema multiagent.

2.4. Sistemes basats en les teories de l'acció situada i la intel·ligència artificial corpòria

[...] that human intelligence is *essentially* embodied; that intelligent bodies are *essentially* situated (embedded in the world); and that the relevant situation (world) is *essentially* human.

J. Haugeland (1996). "Body and World: a Review of What Computers Still Can't Do: a Critique of Artificial Reason".

Una manera diferent de veure la intel·ligència artificial és mitjançant les teories de l'acció situada².

⁽²⁾En anglès, *situated action theories*.

Les teories de l'acció situada argumenten que no s'ha de veure la intel·ligència com un procés de construcció i avaluació de models del món, sinó com un procés menys estructurat d'actuar en el món i respondre als resultats obtinguts. Per tant, es dóna més importància a l'habilitat d'actuar que a la d'explicar aquestes accions; es prima la conducta al raonament.

El fet de subratllar la importància de les accions en si en detriment de com es determina quina és la que s'ha de fer, ha arribat a plantejar la intel·ligència sense representació. Això, evidentment, es pot veure relacionat amb els sistemes que utilitzen xarxes neurals on, com ja s'ha dit, no existeix representació explícita.

Els **sistemes reactius** (també anomenats *sistemes estímul-resposta*) són aquells que no tenen un estat intern i que simplement reaccionen als estímuls que hi ha en l'entorn.

La visió d'actuar en el món i de respondre al que hi succeeix implica donar molta importància a la nostra habilitat per a sentir (*sense*) el que hi ha al nostre voltant i a respondre de manera ràpida i continuada als canvis que s'hi succeeixen. Això porta a subratllar la importància dels cossos i dels sentits que ens situen en l'entorn en lloc de donar-la als processos d'un nivell més alt que raonen sobre aquest entorn.

En aquesta teoria, a diferència del cas de la hipòtesi dels sistemes de símbols físics, la instanciació del sistema de símbols físics és rellevant. Per a fer una acció intel·ligent en el món és necessari un cos físic que permeti a l'agent d'estar completament situat en el món. La intel·ligència s'obté a partir de l'activitat pràctica de captar el món de cada dia. És l'anomenada *intel·ligència artificial corpòria (embodied)*.

Lectura complementària

R. Brooks té un treball titulat precisament així:

R. A. Brooks (1991). "Intelligence without Reason". *Proceedings of the Int. J. Conf. on Artificial Intelligence* (pàg. 569-595).

Tanmateix, no tothom que critica el fet de tenir els símbols com a únics elements significatius del problema està d'acord amb la necessitat de la intel·ligència artificial corpòria. Hi ha qui afirma, com Collins, que més que parlar d'intel·ligència artificial corpòria hauríem de parlar d'inclusió social³. En aquesta interpretació, no és necessària la presència física mitjançant un cos, sinó que el que és imprescindible és la relació amb els agents socials d'aquest entorn. És imprescindible la socialització dels sistemes per aconseguir la seva intel·ligència.

De fet, Collins afirma que el món conceptual ja està "situat":

The odd thing about all of this is that Dreyfus, as a follower of Wittgenstein, ought not to think of the conceptual world as much different from the perceptual world. The conceptual world too is "situated". The later philosophy of Wittgenstein shows that what we take as logically and scientifically compelling is what we do not know how to doubt. What we take as certain is what follows for us, as a matter of course, in the way we live in the world. In the last resort there is no more compelling proof. Even logical syllogisms cannot be proved if we are unwilling just to "see" and act as though they follow in the situations in which we find ourselves. Thinking and acting in the world are but two sides of the same coin.

I més endavant en el mateix treball, Collins diu:

In sum, the shape of the bodies of the members of a social collectivity and the situations in which they find themselves give rise to their form of life. Collectivities whose members have different bodies and encounter different situations develop different forms of life. But given the capacity for linguistic socialisation, an individual can come to share a form of life without having a body or the experience of physical situations which correspond to that form of life. What we don't know is how to make something with the capacity to be socialized in this way.

H.M. Collins. "Embedded or Embodied? A Review of Hubert Dreyfus' *What Computers Still Can't Do*" (1996).

Aplicacions

La importància d'actuar enfront de la de raonar ha influït en les xarxes neurals i en els sistemes basats en agents, i està, a més, en el cor dels atacs als sistemes simbòlics. En lloc d'aquests sistemes es defensen principalment els processos d'aprenentatge i adaptació.

Aquests models han influït també en els treballs d'R. Brooks. En aquests treballs s'argumenta que la intel·ligència artificial s'equivoca començant pels processos de raonament a nivells alts. Brooks proposa començar amb la construcció i avaluació de sistemes petits i simples (que operen a nivell d'insectes) i que permeten d'estudiar els processos mitjançant els quals tant les criatures senzilles com les complexes actuen en el món. Després d'haver desenvolupat sistemes d'aquest tipus, ara es desenvolupen sistemes més complexos.

Cal dir que aquestes idees han influït també en aspectes de l'anomenada *intel·ligència artificial simbòlica* (la que segueix la hipòtesi dels sistemes de símbols físics). Per exemple, en la planificació reactiva⁴. En aquest tipus de planificació s'abandonen les tècniques tradicionals que intenten construir un pla complet per arribar a l'objectiu des d'un estat inicial. Aquests sistemes no

⁽³⁾En anglès, *social embedding*.

Lectura complementària

Collins es refereix al llibre de Dreyfus següent:

H. Dreyfus (1992). *What Computers Still Can't Do*. Cambridge: MIT Press.

Vegeu també

Els problemes de planificació es defineixen en l'apartat 1.1.1 del mòdul "Resolució i problemes de cerca".

⁽⁴⁾En anglès, *reactive planning*.

funcionen adequadament perquè acostumen a sorgir problemes no previstos quan el pla s'executa. Els sistemes de planificació reactiva s'implementen amb un cicle de:

- 1) construcció de plans parcials;
- 2) actuació del pla i
- 3) reavaluació de la situació per a nous plans.

L'èmfasi en l'acció és cada cop més important en la intel·ligència artificial. D'aquí el ressorgiment de la robòtica i els problemes de percepció. Això explica també l'interès que ha tingut Internet: els agents per a web (o *softbots*) que fan tasques útils a la xarxa Internet.

2.5. Mètodes i tècniques de la intel·ligència artificial

Tot i que hi ha diferents punts de vista, i els objectius del camp a llarg termini poden ser ambiciosos, la intel·ligència artificial es dedica a la construcció de programes. Per a construir-los s'han desenvolupat un conjunt de mètodes i tècniques que són els elements bàsics per a construir sistemes més complexos.

Un esquema dels mètodes i tècniques de la intel·ligència artificial és aquest:

- Resolució de problemes i cerca
- Sistemes basats en el coneixement
- Aprenentatge
- Intel·ligència artificial distribuïda

Aquests temes constitueixen el que se'n pot dir el nucli central de la intel·ligència artificial. Aquests elements s'apliquen a la resolució de problemes en molts camps diferents (com ara la biologia, l'economia, l'educació o la història). Entre els camps als quals s'apliquen aquestes tècniques n'hi ha uns que, de fet, es pot dir que formen part de la mateixa intel·ligència artificial perquè influeixen directament en el seu desenvolupament. Aquests camps són:

- El llenguatge natural.
- La visió.
- La robòtica.
- El reconeixement de la parla.

No tothom està d'acord amb aquesta separació a dos nivells. Winston afirma:

En ocasiones, se escucha la frase IA central utilizada por gente que considera al *lenguaje*, la *visión* y la *robótica* como corrientes un tanto separadas de la corriente principal de la inteligencia artificial. Sin embargo, por la forma en que los progresos obtenidos en lenguaje, visión y robótica pueden influir y de hecho han influido en los trabajos acerca del razonamiento, cualquier separación de este tipo no parece ser muy aconsejable.

Winston (1994). *Inteligencia Artificial*.

En aquesta assignatura s'estudiaran alguns dels mètodes i tècniques del primer grup. En concret, ens centrarem en els temes de resolució de problemes i cerca, i en els sistemes basats en el coneixement. També es farà una introducció a les diferents formes de supervisió dels algorismes d'aprenentatge computacional: supervisat, no supervisat i per reforç.

És a l'assignatura d'*Aprenentatge computacional* on s'estudiarà amb més detall els diferents algorismes d'aprenentatge i també s'introduirà la intel·ligència artificial distribuïda a partir dels sistemes multiagents.

Entre les aplicacions que hi ha actualment al voltant de les tècniques d'aprenentatge podem subratllar la prospecció de dades⁵, també denotada per descobriment de coneixement de bases de dades⁶. Això correspon a extreure informació potencialment útil, que abans no es coneixia, però que estava implícita en unes dades.

⁽⁵⁾En anglès, *data mining*.

⁽⁶⁾En anglès, *knowledge discovery in databases*.

La intel·ligència artificial distribuïda correspon a situacions en què hi ha diversos sistemes que interaccionen per a resoldre un problema comú. Dintre del camp hi ha dues àrees principals: la resolució de problemes de manera distribuïda⁷ i els sistemes multiagents⁸. En la primera àrea s'estudia com dividir una tasca en mòduls que cooperen i comparteixen informació sobre un problema i sobre les solucions. Tanmateix, en aquest cas tota forma d'interacció entre mòduls (les estratègies de cooperació i coordinació) està prefixada i incorporada en el sistema. En la segona àrea, s'estudia la construcció i la conducta d'una col·lecció d'agents autònoms que intenten resoldre un problema. Aquests sistemes poden ser heterogenis, això és, amb capacitats de raonament diferents.

⁽⁷⁾En anglès, *distributed problem solving*.

⁽⁸⁾En anglès, *multiagent systems*.

3. Algunes aplicacions

Fins ara s'han desenvolupat molts sistemes que utilitzen les tècniques desenvolupades en el camp de la intel·ligència artificial. Aquests sistemes han estat construïts per a ser utilitzats en entorns molt diferents (d'entorns industrials a objectes d'ús quotidià, des d'aplicacions mèdiques a eines per a ajudar a l'aprenentatge, per exemple, de matèries relacionades amb la història). En aquest apartat esmentem algunes aplicacions. Primer, repassem alguns dels sistemes que van ser innovadors en el moment de la seva construcció i acabem amb alguns sistemes més recents (entre parèntesis s'inclouran els investigadors principals de cada sistema i l'any de publicació del resultat):

1) Logic Theorist (Newell, Shaw, Simon, 1957). Era un demostrador automàtic de teoremes. Això és, demostrava un teorema a partir d'un conjunt d'axiomes donat. Entre el que va demostrar hi ha teoremes del segon capítol del llibre de Russell i Whitehead titulat *Principia Mathematica*. A més, d'un d'aquests teoremes en va fer una demostració més curta que la donada en el llibre. Com ja hem dit abans, el Logic Theorist va ser presentat en l'estada de Dartmouth de l'estiu del 1956.

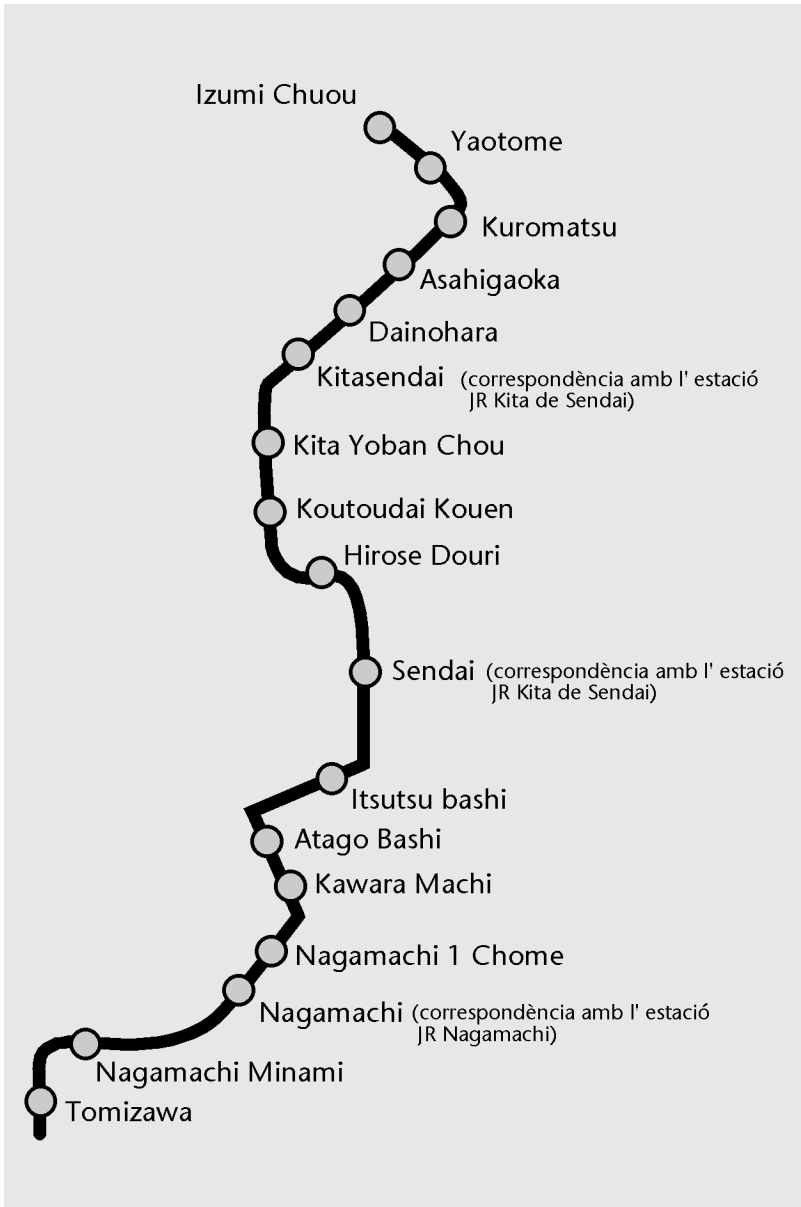
2) Mycin (Shortliffe i Buchanan, 1975). Era un sistema expert dissenyat per a ajudar en la recomanació de teràpies apropiades per a pacients amb infeccions bacterianes. El sistema utilitzava regles per a raonar que, per tal de modelitzar la incertesa, incorporaven els anomenats *factors de certesa*.

3) AM (Lenat, 1977). Era un programa experimental per a construir teories de manera automàtica en el camp de les matemàtiques. A partir d'un conjunt de conceptes elementals de la teoria de conjunts i de l'aritmètica (i de la definició d'aquests conceptes), AM produïa nous conceptes i conjectures a partir d'uns mecanismes generadors (operadors). A més, mitjançant un conjunt de regles seleccionava d'entre els conceptes generats aquells que creia que podien ser interessants. Entre els conceptes que va generar, hi ha el de nombre primer. Era un dels primers sistemes de descobriment automàtic.

4) Prospector (Duda, 1979). Era un sistema expert construït per a ajudar en la localització de dipòsits de minerals com el coure i l'urani. Va ser un dels sistemes que va contribuir al renom de la intel·ligència artificial cap allà als anys vuitanta en recomanar l'excavació en un lloc on es va trobar un dipòsit de molibdè molt gran.

5) Metro de Sendai (Yasunobu, 1987). El sistema de control per a la conducció automàtica del metro de la ciutat japonesa de Sendai va ser desenvolupat per Seiji Yasunobu del laboratori de desenvolupament de sistemes d'Hitachi. El sistema està construït mitjançant regles difuses. Actualment hi ha sistemes basats en regles difuses en molts aparells d'ús quotidià (rentadores, càmeres de fotografar, aparells d'aire condicionat, etc.).

Figura 1



Plànol del metro de la ciutat de Sendai.

6) EQP (McCune, 1997). El sistema EQP va demostrar l'any 1996 l'anomenat *problema de Robbins*. Aquest problema, que va ser plantejat per H. Robbins els anys trenta, és molt conegut en l'àlgebra booleana i fins ara era un problema obert (que no s'havia pogut resoldre). El sistema EQP és un programa de demostració automàtica per a lògica equacional. La solució va ser trobada de manera automàtica entre setembre i octubre del 1996.

Lectura complementària

Trobareu més informació sobre el sistema de control del metro de Sendai i sobre Deep Blue, respectivament, a les adreces següents:

<http://funny.esys.tsukuba.ac.jp/~yasunobu/aaaeg.htm>

<http://www.research.ibm.com/deep-blue/>

7) Deep Blue (IBM, 1997). Per primer cop un ordinador jugador d'escacs va guanyar al campió humà mundial, G. Kasparov a Nova York el mes de maig de l'any 1997. El programa fa servir algorismes de cerca paral·lela basats en els proposats per C. Shannon, a més de maquinari dedicat. Deep Blue té coneixement del domini incorporat en la funció que utilitza per a avaluar els diferents taulers que plantegen les diferents opcions (funció construïda per M. Campbell amb la col·laboració dels jugadors d'escacs J. Benjamin i M. Illescas). A més, Deep Blue incorpora bases de dades de finals del joc que permeten jugar un joc perfecte quan s'està en aquelles posicions.

8) DARPA Grand Challenge (Stanford University, 2005). El DARPA Grand Challenge consisteix en una carrera de vehicles autònoms que han d'arribar d'un punt dels Estats Units d'Amèrica a un altre punt sense intervenció humana i disposant únicament d'un llistat de punts intermedis entre el principi i el final del circuit. L'any 2005, per primera vegada, cinc vehicles autònoms van aconseguir completar els 212 km de què consta el circuit, ocupant el primer lloc un vehicle de l'Stanford University, en un temps de 6 h 53 min.

9) DARPA Urban Challenge (CMU, 2007). A diferència del DARPA Grand Challenge de 2005, aquest va tenir lloc en una àrea urbana, en un circuit de 96 km on els vehicles havien de complir la normativa de trànsit, a més d'evitar topar amb altres vehicles de la cursa, tant vehicles autònoms com conductors professionals. En les edicions anteriors, els vehicles no trobaven altres vehicles en el recorregut, malgrat que les condicions físiques del terreny eren més desafians. El vehicle autònom guanyador va ser dissenyat per la Carnegie Mellon University (CMU) i va completar el circuit en un temps de 4 h 10 min.

10) Watson (IBM, 2011). L'any 2011, un sistema informàtic dissenyat per IBM, capaç de respondre preguntes formulades en llenguatge natural, va guanyar als dos millors participants humans que havia tingut un concurs de televisió anomenat Jeopardy! Es tracta d'un concurs amb preguntes de diverses temàtiques com ara història, idiomes, literatura, cultura popular, belles arts, ciències, geografia i esports. El programa informàtic tenia accés a dos-cents milions de pàgines de contingut estructurat i no estructurat però no disposava de connexió a internet durant el joc. El projecte va ser desenvolupat per un equip de recerca amb la tecnologia DeepQA⁹ d'IBM.

⁽⁹⁾En anglès, QA fa referència a *Question Answering* (resposta de preguntes).

11) AlexNet (Krizhevsky, 2012). El disseny i implementació amb GPU d'una xarxa neuronal convolucional, formada per vuit capes (cinc convolucionals i tres totalment connectades), va suposar una millora molt significativa en una competició de reconeixement visual d'imatges anomenada ImageNet, passant d'un error del 26,1 % a un error del 15,3 %. La tasca consisteix en donar les cinc categories més probables d'una imatge entre les mil possibles categories que hi ha definides. A partir de 2015, amb xarxes cada cop més profundes, es va aconseguir un error inferior a la taxa d'error humana, la qual està estimada en un 5 %.

12) AlphaGo (Google DeepMind, 2015). Per primer cop, l'octubre de 2015, un programa informàtic derrotava un jugador professional de Go, un joc d'estratègia de tauler per a dos jugadors molt popular a l'Àsia Oriental que es juga en un tauler 19x19, fent que la complexitat sigui significativament superior a la del joc d'escacs. El programa utilitza un algorisme de cerca en arbre per a trobar els moviments basant-se en el coneixement prèviament après amb una xarxa neuronal a partir tant de les partides jugades pels humans com de les partides jugades pel mateix ordinador.

Lectura complementària

AlphaGo va tenir un ressò mediàtic important, amb la producció inclús d'un documental que va rebre diversos premis en festivals de cinema.

[https://
www.alphagomovie.com](https://www.alphagomovie.com)

4. Algunes característiques dels sistemes intel·ligents

Fins ara s'han construït molts sistemes intel·ligents que s'apliquen en entorns molt diferents. De tota manera, tot i la diversitat de sistemes i de camps d'aplicacions, hi ha alguns elements comuns en tots aquests programes. A continuació es descriuen aquests elements. És important subratllar que no tots els sistemes els apliquen tots, i que la importància que es dona a cada element depèn molt del sistema.

Les característiques d'aquests sistemes són les següents:

- 1) **Usen informació simbòlica.** La informació que han de tractar els sistemes és sobretot simbòlica. Han de raonar sobre coses que passen (fets), conceptes abstractes i han de treure conclusions.
- 2) **Fan servir descripcions del domini de l'aplicació.** Per a poder trobar una solució, els sistemes han de conèixer l'entorn del problema. Han de saber què hi ha, com hi poden actuar, com evoluciona l'entorn.
- 3) **Empren dades incompletes, inexactes o amb conflicte.** La informació que han de tractar els sistemes no està en les condicions ideals que es voldria. Sovint s'ha de tractar el fet que manca informació, o que les dades rebudes estan en conflicte (entre aquestes o amb el coneixement que el sistema ja té incorporat). També hi ha problemes d'incertesa (valors aproximats, conceptes difusos, etc.).
- 4) **Utilitzen mètodes heurístics.** Els programes fan allò que se'ls diu, però en el cas de moltes aplicacions d'intel·ligència artificial el que tenim implementat són indicacions de com s'ha de trobar la solució més que no pas descripcions detallades amb tots els passos a seguir. Els mètodes heurístics ens donen una idea de com hem de cercar la solució, però això no ens garanteix l'èxit. Sovint no hi ha la millor solució, però se'n troba una de prou bona per a operar.
- 5) **Són adaptatius.** Quan l'entorn en el qual està un sistema canvia, el sistema també ha de canviar per tal que el seu comportament continuï essent l'adequat.

Vegeu també

Vegeu algunes aplicacions de la intel·ligència artificial en l'apartat 3 d'aquest mòdul.

Per exemple,...

... molts dels sistemes basats en el coneixement (vegeu el mòdul "Sistemes basats en el coneixement" d'aquesta assignatura) no són adaptatius; alguns sistemes reactius no utilitzen informació simbòlica.

Exemple de descripció del domini d'aplicació

En un sistema de diagnòsi per a un determinat tipus d'aparells, s'hauran de conèixer les relacions entre els símptomes i les causes, a més s'haurà de saber com es comporten els components de l'aparell quan tot funciona correctament.

Activitats

1. Llegiu l'article d'A. Turing publicat l'any 1950 a la revista *Mind* (núm. 59, pàg. 433-60), i considereu les prediccions que va fer per a l'any 2000.
2. Considereu les diferents maneres d'entendre la intel·ligència artificial. Quina us sembla més encertada?
3. Creieu que és cert, tal com afirma Simon, que la intel·ligència artificial és una ciència experimental?
4. Quina de les quatre definicions alternatives sobre la IA considerant conducta enfront de raonament i racionalitat enfront de comparació amb el comportament de les persones, us sembla més adequada?
5. Considereu els problemes que es presenten a continuació. El tres en ratlla i el parxís són programes de jocs, i la manera com tractar-los es descriu en el mòdul "Resolució de problemes i cerca" d'aquesta assignatura. El mateix passa amb el problema de la integració d'expressions. El cas dels circuits és considerat com un exemple del raonament basat en models. En el mòdul "Sistemes basats en el coneixement" d'aquesta assignatura es dedica un apartat a aquest tema. De tota manera, abans de passar als mòduls podeu respondre a les preguntes següents.
 - a) Considereu el joc del tres en ratlla. Com ho faríeu per a triar la millor jugada? I en el cas del parxís?
 - b) Considereu el problema de la integració d'expressions (per exemple, quina és la integral de $x^2 + 3e^x$). Com resoldríeu aquest problema?
 - c) Disposeu d'una fàbrica de construcció de circuits i voleu construir un sistema que comprovi el funcionament de cada circuit i quan no funcioni determinar què és el que falla. Com tractaríeu aquest problema?

Bibliografia

Bibliografia bàsica

Haton, J. P.; Haton, M. C. (1993). *L'intelligence Artificielle*. París: Presses Universitaires de France (traducció catalana de Tibidabo edicions).

Luger, G. F. (1995). *Computation & Intelligence*. Menlo Park: MIT Press.

Bibliografia complementària

Brooks, R. A. (1991). . "Intelligence without Reason". *Proceedings of the Int. J. Conf. on Artificial Intelligence* (pàg. 569-595).

Collins, H. M. (1996). . "Embedded or embodied? A review of Hubert Dreyfus' 'What Computers Still Can't Do'". *Artificial Intelligence* (núm. 80, pàg. 99-117).

Dreyfus, H. L. (1992). *What Computers Still Can't Do: A Critique of Artificial Reason*. Cambridge: MIT Press.

Haugeland, J. (1996). . "Body and World: a Review of What Computers Still Can't Do: A Critique of Artificial Reason (Huber L. Dreyfus)". *Artificial Intelligence* (núm. 80, pàg. 119-128).

Newell, A.; Simon, H. A. (1976). "Computer Science as Empirical Inquiry: Symbols and Search". *Communications of the ACM* (vol. 3, núm. 19, pàg. 113-126).

Russell, S.; Norvig, P. (1995). *Artificial Intelligence: A Modern Approach*. Englewood Cliffs: Prentice Hall.

Simon, H. A. (1995). "Artificial Intelligence: an Empirical Science". *Artificial Intelligence* (núm. 77, pàg. 95-127).