
Qué es la inteligencia artificial

PID_00267999

Vicenç Torra i Reventós

Tiempo mínimo de dedicación recomendado: 2 horas



Vicenç Torra i Reventós

La revisión de este recurso de aprendizaje UOC ha sido coordinada por el profesor: Carles Ventura Royo (2019)

Tercera edición: septiembre 2019
© Vicenç Torra i Reventós
Todos los derechos reservados
© de esta edición, FUOC, 2019
Av. Tibidabo, 39-43, 08035 Barcelona
Realización editorial: FUOC

Ninguna parte de esta publicación, incluido el diseño general y la cubierta, puede ser copiada, reproducida, almacenada o transmitida de ninguna forma, ni por ningún medio, sea este eléctrico, químico, mecánico, óptico, grabación, fotocopia, o cualquier otro, sin la previa autorización escrita de los titulares de los derechos.

Índice

Introducción.....	5
Objetivos.....	6
1. Breve repaso histórico de la inteligencia artificial.....	7
2. Qué es la inteligencia artificial: definiciones y puntos de vista.....	11
2.1. La hipótesis del sistema de símbolos físicos	13
2.2. Sistemas basados en modelos biológicos	15
2.3. Sistemas emergentes	16
2.4. Sistemas basados en las teorías de la acción situada y la inteligencia artificial corpórea	17
2.5. Métodos y técnicas de la inteligencia artificial	19
3. Algunas aplicaciones.....	21
4. Algunas características de los sistemas inteligentes.....	25
Actividades.....	27
Bibliografía.....	28

Introducción

Este módulo presenta una visión general de la inteligencia artificial e introduce la asignatura. Con el fin de proporcionar un aspecto global del área y de las aplicaciones existentes, el módulo se ha dividido en cuatro apartados.

Empezaremos con un apartado dedicado a realizar un breve repaso de la historia del área a partir de algunos de los acontecimientos más importantes.

A continuación, se dedica un apartado a ver qué es la inteligencia artificial. Dado que la opinión de los investigadores no coincide a la hora de definirla ni de decir cuáles son los objetivos principales, presentamos varios puntos de vista. La descripción nos permite encuadrar la asignatura dentro del cuerpo general de la inteligencia artificial.

El tercer apartado está dedicado a describir algunas de las aplicaciones existentes, como demostradores de teorías o sistemas expertos.

El módulo acaba con un apartado que incluye un repaso de las características que comparten los sistemas que utilizan técnicas de inteligencia artificial.

Objetivos

Este módulo didáctico pretende alcanzar los objetivos siguientes:

- 1.** Conocer los diferentes puntos de vista que existen sobre la inteligencia artificial.
- 2.** Saber enmarcar el temario de la asignatura dentro del área.
- 3.** Tener una idea general de la historia de la inteligencia artificial.
- 4.** Conocer algunos ejemplos prácticos en los que se aplican las técnicas de inteligencia artificial a problemas reales.

1. Breve repaso histórico de la inteligencia artificial

El término *inteligencia artificial* (IA) fue adoptado durante el verano de 1956 en Dartmouth en un encuentro que reunió a investigadores interesados en los temas de la inteligencia, las redes neuronales y la teoría de autómatas. Este encuentro fue organizado por J. McCarthy (que se encontraba en Dartmouth), C. Shannon, M. Minsky y N. Rochester y asistieron a ella T. More (Princeton), A. Samuel (IBM), R. Solomonoff y O. Selfridge (MIT) y A. Newell, H. Simon (Carnegie Tech, actualmente Carnegie Mellon University). La reunión constituyó lo que después sería el núcleo central de la inteligencia artificial.

Notación

Abreviamos el término *inteligencia artificial* con la sigla IA.

El encuentro fue la consecuencia de los primeros trabajos que se estaban realizando en este campo. Por ejemplo, Newell y Simon presentaron su programa de razonamiento automático Logic Theorist (que causó mucha sensación). De hecho, actualmente se considera que la inteligencia artificial nace con un trabajo desarrollado por W. McCulloch y M. Pitts de 1943 en el que se proponía un modelo de neuronas artificiales basado en tres fuentes: conocimientos sobre la fisiología y el funcionamiento de las neuronas del cerebro; la lógica proposicional de Russell y Whitehead, además de la teoría de la computación de Turing. De unos años más tarde fue el primer ordenador neuronal (el SNARC). Fue construido en torno al año 1951 por dos estudiantes de Princeton: M. Minsky y D. Edmons. Más o menos de la misma época fueron los primeros programas para jugar al ajedrez de Shannon y Turing.

Aunque éstos son los trabajos con los que se inicia la inteligencia artificial, existen muchos otros anteriores que han influido en la evolución de esta área. Algunos de estos trabajos han aparecido en los campos de la filosofía (uno de los primeros intentos de formalizar el razonamiento son los silogismos del filósofo griego Aristóteles), las matemáticas (la teoría de la decisión y la teoría de la probabilidad se aplican en muchos de los sistemas actuales) o la psicología (que con la inteligencia artificial han definido el campo de la ciencia cognitiva).

En los años siguientes a la reunión de Dartmouth se consiguieron muchos progresos. Se construyeron programas que resolvían problemas muy diferentes. Por ejemplo, estudiantes de Minsky implementaron a finales de la década de los sesenta los programas *Analogy*, que resolvían problemas geométricos como los que aparecen en los tests de inteligencia, y *Student*, que daba solución a problemas de álgebra redactados en inglés. Newell y Simon desarrollaron el *General Problem Solver* (GPS), que imitaba el razonamiento humano. Samuel escribió programas para jugar a los *checkers* (juego parecido a las damas) que aprendían a jugar mejor que el autor. McCarthy, que se había marchado al MIT, implementó el año 1958 el lenguaje Lisp y escribió durante aquel mismo

año un trabajo titulado *Programs with Common Sense* en el que se describe un programa hipotético que constituye el primer sistema completo de inteligencia artificial.

Esta serie de éxitos queda truncada a mediados de la década de los sesenta y las previsiones, demasiado optimistas, de aquellos años se frustran. Los sistemas implementados hasta aquel momento funcionaban en dominios limitados (conocidos como micromundos¹). La transformación para adaptarse a entornos reales no fue (no ha sido) trivial como se esperaba. Esto se debió, fundamentalmente, según Russell y Norvig, a los tres factores siguientes:

1) Algunos de los sistemas no disponían de conocimiento del entorno de la aplicación o éste resultaba muy pequeño y únicamente consistía en algunas manipulaciones sintácticas simples.

ELIZA: un sistema de manipulaciones sintácticas simples

El programa de diálogos ELIZA de Weizenbaum quería simular a un psicólogo que conversaba con el paciente. Su funcionamiento se basaba en el reconocimiento de estructuras en las frases del paciente y sobre la base de unos patrones construía unas nuevas frases. Disponía de reglas como las dos que ofrecemos a continuación (donde para cada regla existe el patrón y una posible respuesta o más):

```
(((* x) recuerdo (* y))
  (¿piensas a menudo en (* y) ?)
(((* x) he soñado (* y))
  (¿seguro-- (* y) ?)
  (¿y ahora por qué me hablas de sueños?)
  (¿has soñado con (* y) antes?))
```

Con estas reglas, cuando alguien escribe “ahora todavía recuerdo el viaje a París”, ELIZA responde “¿piensas a menudo en el viaje a París?” porque reconoce la estructura “recuerdo” y asocia lo que sigue a esta palabra a la variable *y* y después construye una nueva frase con el contenido de esta *y*. Lo mismo hará con la segunda regla. En este caso, hay tres posibles respuestas.

2) Muchos de los problemas que se intentaban resolver eran problemas intratables (NP-complejos). Mientras la cantidad de conocimiento resultaba pequeña, se podían resolver; pero cuando esta era mayor, los problemas resultaban irresolubles.

Ejemplo de problema intratable

Un problema que considera permutaciones de objetos al pasar de 8 a 10 objetos tarda unas 100 veces más. ¡Si pasamos a 20 objetos es 6×10^{13} veces más!

3) Algunas de las estructuras básicas que se utilizaban para generar una conducta inteligente tenían grandes limitaciones.

Limitaciones de los perceptrones

Los perceptrones (un tipo de red neuronal) desarrollados por Minsky y Papert no podían aprender la función *or* exclusiva (*xor*). Esto es, no existe ningún perceptrón (ninguna asignación de pesos a las conexiones) con dos unidades de entrada y una de salida que se comporte como la función *or* exclusiva. Es decir, que cuando las dos entradas son cero,

⁽¹⁾En inglés, *microworlds*.

Lectura complementaria

Podéis ver las principales dificultades de adaptación de los sistemas inteligentes a entornos reales en el primer capítulo de la obra siguiente:

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

Algoritmos NP

De manera informal definimos un algoritmo NP como un algoritmo que necesita un tiempo que crece exponencialmente con la medida de los datos de entrada.

el resultado sea cero; que si las dos son 1, también dé 0; y que cuando una de las dos es 1, el resultado sea 1.

Entrada	Entrada	Función <i>xor</i>
0	0	0
0	1	1
1	0	1
1	1	0

Por lo tanto, los perceptrones no servían para cualquier tipo de problema. En particular, no sirven para los problemas de clasificación que no son linealmente separables.

Hasta aquel momento, la resolución de problemas se basaba en un mecanismo de búsqueda de propósito general que intentaba encadenar pasos elementales de razonamiento para encontrar la solución completa. El coste de este mecanismo resultaba muy elevado; para reducirlo se desarrollaron los primeros algoritmos de búsqueda. Por ejemplo, Dijkstra durante el año 1959 diseñó el método de coste uniforme, Newell y Ernst en el año 1965 desarrollaron el concepto de búsqueda heurística y Hart, Nilsson y Raphael, el algoritmo A*. En esta misma época, en relación con los programas de juegos, se definía la búsqueda alfa-beta. De hecho, fue ideada por McCarthy en el año 1956 y utilizada más tarde (en 1958) por Newell.

La importancia de considerar el conocimiento del dominio de la aplicación y su estructuración a fin de que su acceso fuera eficiente e hiciera posible la construcción de sistemas para entornos reales llevó al estudio de los métodos de representación del conocimiento. Entre otros, se definieron las redes semánticas (definidas a principios de los años sesenta por Quillian) y los marcos (definidos por Minsky en 1975). En este mismo período se empezaron a utilizar algunos tipos de lógica para representar el conocimiento.

De manera paralela, durante estos años continúan los trabajos, en la línea del sistema para jugar a los *checkers* desarrollado por Samuel, orientados a incorporar algún tipo de aprendizaje. E.B. Hunt, J. Marin y P.T. Stone construyen durante el año 1966 los árboles de decisión (para clasificación), ya ideados por Shannon en 1949, y Quinlan (1979) presenta el método ID3 para construir estos árboles. Por otra parte, P. Winston (1970) desarrolla su programa para aprender descripciones de objetos complejos y T. Mitchell (1977) desarrolla el llamado *espai de versions*. Más tarde, a mediados de la década de los ochenta, el redescubrimiento del método de aprendizaje para redes neuronales denominado *retropropagació* (“backpropagation”) lleva al resurgimiento de este campo.

Ved también

Podéis ver los algoritmos de coste uniforme y A* en los subapartados 4.1 y 4.3 del módulo “Resolución de problemas y búsqueda” de esta asignatura.

Ved también

El algoritmo de retropropagación y los modelos basados en redes neuronales se introducen en el módulo 5 “Introducción al aprendizaje computacional” de esta asignatura y se verán con más detalle en la asignatura *Aprendizaje computacional*.

La construcción de aplicaciones para entornos reales ha obligado a tratar aspectos como la incertidumbre o la imprecisión (necesarios también para resolver los juegos de azar). Se han aplicado métodos probabilísticos (teoría de la probabilidad o redes probabilísticas) y se han desarrollado otros formalismos como los conjuntos difusos (definidos por L. Zadeh en 1965) o la teoría de Dempster-Shafer (original de A. Dempster el año 1968 con aportaciones de G. Shafer del año 1976).

Con estos desarrollos se construyeron a partir de los años ochenta los primeros sistemas comerciales (principalmente sistemas expertos).

Hoy en día los problemas de investigación en el campo de la inteligencia artificial van orientados hacia la creación de sistemas cooperativos basados en agentes, incluyendo sistemas para la recuperación y gestión de datos, visión por computador y procesamiento de lenguaje natural, entre otros.

Ved también

Algunos de los aspectos relacionados con la incertidumbre y la imprecisión se describen en el módulo "Incertidumbre y razonamiento aproximado" de esta asignatura.

2. Qué es la inteligencia artificial: definiciones y puntos de vista

A diferencia de lo que ocurre en otras áreas, en la inteligencia artificial no existe una única definición, sino que hay muchas correspondientes a puntos de vista diferentes. En este apartado empezaremos con un pequeño repaso de algunas de las definiciones existentes. Después se verán las hipótesis sobre las que se han desarrollado los programas.

Construcción de programas

La inteligencia artificial, como campo de la informática que es, está dedicada a la construcción de programas.

Utilizando una frase de Simon podemos decir que:

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

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

Conducta frente a razonamiento

De todos modos, los objetivos que persiguen estos programas nos llevarán a diferentes definiciones. En su libro de inteligencia artificial, Russell y Norvig clasifican los objetivos de acuerdo con dos dimensiones:

1) Por una parte, el objetivo de los programas puede ser obtener una conducta determinada (nos interesa el resultado que se obtiene –el comportamiento–), o bien conseguir una manera de razonar determinada (más que el resultado lo que nos interesa es que el razonamiento de una materia se realice de una manera determinada). Este par conducta/razonamiento constituye la primera dimensión.

2) Por otra, se exigirá que los programas sean correctos. Esta corrección se puede medir comparando el rendimiento con el de las personas o bien comparándolo con un concepto ideal de inteligencia. Este ideal lo denominamos *racionalidad*. La forma de medir la corrección define la segunda dimensión.

De acuerdo con estas dos dimensiones existen cuatro visiones posibles para los objetivos y, por lo tanto, cuatro definiciones posibles para el término *inteligencia artificial*. Estas alternativas han sido consideradas a lo largo de la historia del campo y son éstas:

Lectura complementaria

Encontraréis la clasificación de los objetivos de la IA según Russell y Norvig en el primer capítulo de la obra siguiente:

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

a) Actuar como las personas: el test de Turing

Este enfoque está basado en la propuesta de A. Turing que define la inteligencia de manera operacional: una conducta es inteligente cuando tiene suficiente nivel para confundir a un interlocutor humano. La descripción del test aparece en un artículo del año 1950 de la revista *Mind* (núm. 59, págs. 433-460).

b) Razonar como las personas: el modelo cognoscitivo

En este enfoque no interesa únicamente el resultado, sino también cómo se obtiene. Se quiere que el proceso de razonamiento resulte parecido al de las personas. El primer sistema con estos objetivos fue el General Problem Solver de Newell y Simon del año 1961. El campo de la ciencia cognitiva utiliza este punto de vista.

c) Razonar racionalmente: las leyes del pensamiento

Este punto de vista se basa en plantear qué significa pensar correctamente o, lo que resulta equivalente, saber cuándo se puede deducir de manera lógica un hecho nuevo a partir del conocimiento del que se dispone. Los sistemas que utilizan algún tipo de lógica se construyen bajo este punto de vista.

d) Actuar racionalmente: el agente racional

Actualmente el término *agente* se encuentra muy difundido. Sobre qué es un agente, el Artificial Intelligence FAQ dice:

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 resulta importante que las conclusiones a las que llegue un sistema se obtengan a partir de deducciones lógicas (inferencias correctas) del conocimiento, sino que se consigan los objetivos (a partir de las propias creencias –de aquello que se cree que es cierto–). Además, la evaluación de estos objetivos (el aspecto de la corrección) se realiza utilizando un concepto general de racionalidad que se puede definir de manera explícita.

Cuando comparamos este último punto de vista con el de las leyes del pensamiento, debemos subrayar que ahora la inteligencia artificial no se encuentra limitada a la inferencia lógica. Podemos utilizar otros mecanismos si son adecuados para conseguir los resultados.

Ciencia e ingeniería

Lectura complementaria

Se puede encontrar una copia del trabajo de Turing mencionado en la dirección siguiente:

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

Ejemplo de actuación racional

En un programa de juegos, la racionalidad sería que el programa de juegos ganara siempre con independencia de si las personas ganan o no.

Se ha comentado anteriormente que la inteligencia artificial se dedica a la construcción de programas. Desde este punto de vista, y más cuando tenemos en cuenta que no se escriben sólo programas, sino que también se diseñan mecanismos como robots, el campo se ha de ver como una ingeniería. Sin embargo, no es solamente ingeniería, sino también una ciencia. Esto se debe al hecho de que además de desarrollar sistemas nuevos, también se estudia y se obtiene nuevo conocimiento sobre los sistemas inteligentes.

Al mismo tiempo, como subraya Simon, es una ciencia experimental dado que se experimenta cuando diseñamos un sistema y observamos cómo cambia en la conducta si cambia el diseño. También se experimenta cuando se utiliza la inteligencia artificial para estudiar la inteligencia natural.

La dualidad del campo explica la existencia en la literatura tanto de trabajos orientados hacia las cuestiones prácticas, como de otros orientados a los aspectos formales. Éstos últimos resultan necesarios para construir y explicar el comportamiento de los primeros.

Veamos ahora las hipótesis que son el fundamento de la mayoría de los sistemas en inteligencia artificial.

2.1. La hipótesis del sistema de símbolos físicos

En el centro de la mayoría de los trabajos en inteligencia artificial se encuentra la llamada *hipótesis del sistema de símbolos físicos* de A. Newell y 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 símbolos físicos** dispone de un conjunto de entidades, los símbolos, que son patrones físicos. Además, existe un conjunto de procesos (operaciones) que operan sobre expresiones para producir expresiones nuevas.

Dado que los sistemas cuentan con recursos limitados, Newell y Simon consideran que se deben realizar búsquedas heurísticas para resolver los problemas. Esta hipótesis es la llamada *hipótesis de la búsqueda heurística*:

Lectura complementaria

Encontraréis la opinión de Simon sobre la IA como ingeniería en la obra siguiente:

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

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).

Considerando las dos hipótesis, tenemos que la actividad inteligente se consigue mediante el uso de:

- **Símbolos:** representan los elementos significativos de un problema.
- **Operaciones:** que operan sobre los símbolos y permiten generar las soluciones.
- **Búsqueda:** para seleccionar una solución entre las soluciones posibles que se pueden generar a partir de las operaciones.

Las dos hipótesis anteriores justifican los esfuerzos para construir máquinas inteligentes y hacen explícitas las suposiciones que forman la base de la inteligencia artificial. En particular, permiten distinguir entre los patrones de símbolos y el medio que se utiliza para implementarlos. Si la inteligencia sólo se obtiene de la estructura de un sistema simbólico, entonces cualquier medio que implemente de manera adecuada estos patrones y los procesos que operan en éste podrá conseguir la inteligencia con independencia de si el sistema está formado de neuronas o de circuitos lógicos.

La hipótesis del sistema de símbolos físicos explica los dos elementos más importantes de la investigación y el desarrollo de aplicaciones en IA:

- 1) Definir las estructuras simbólicas y las operaciones necesarias para resolver problemas.
- 2) Desarrollar estrategias para buscar, de manera correcta y eficiente, las soluciones potenciales generadas por estas estructuras y operaciones.

Éstos son los aspectos interrelacionados de búsqueda y representación del conocimiento. Estos dos elementos constituyen el núcleo del área.

La opinión de Newell y Simon es, sin embargo, discutida por muchos autores dentro de la inteligencia artificial. Algunos han construido sistemas que se basan en hipótesis diferentes. Podemos subrayar tres alternativas distintas basadas en tres modelos diferentes de la inteligencia. A continuación se presentan las tres alternativas de manera independiente aunque, de hecho, hay fuertes relaciones entre sí: los sistemas basados en modelos biológicos, los sistemas emergentes y los sistemas basados en las teorías de la acción situada y la inteligencia artificial corpórea.

Ved también

El tema de la búsqueda se ve en el módulo “Resolución de problemas y búsqueda” y la representación se ve en el módulo “Sistemas basados en el conocimiento” de esta asignatura.

2.2. Sistemas basados en modelos biológicos

Una alternativa a los sistemas de símbolos físicos la constituyen las redes neuronales, que han sido diseñadas bajo la inspiración de modelos biológicos.

Una **red neuronal** está formada por “neuronas” (unidades de procesamiento) y unas conexiones entre sí que tienen asociadas unos pesos.

Ved también

Los modelos basados en redes neuronales se introducen en el módulo 5 “Introducción al aprendizaje computacional” de esta asignatura y se tratan con más detalle en la asignatura *Aprendizaje computacional*.

Estas redes representan de manera implícita el mundo mediante los pesos asociados a cada conexión. Por lo tanto, no se utilizan símbolos. Además, a causa de que el conocimiento se encuentra distribuido por toda la estructura de la red resulta muy difícil, si no imposible, aislar los conceptos individuales en regiones específicas de la red. De hecho, una porción de la red se puede utilizar para representar diferentes conceptos.

Las redes desplazan el énfasis de la inteligencia artificial fuera de los problemas de la representación simbólica y de las estrategias de inferencia sólida a problemas de aprendizaje y adaptación. Las redes neuronales son mecanismos que se adaptan al mundo (adoptan su forma por la interacción con el mundo mediante técnicas de aprendizaje que modifican los pesos de las conexiones), y no necesitan que el mundo sea refundido en un modelo simbólico explícito.

Por todo ello, las redes neuronales se pueden ver (como mínimo) como un contraejemplo de la condición de necesidad de la hipótesis de los sistemas de símbolos físicos.

Aunque la mayoría de los autores considera que las redes vulneran la hipótesis de Newell y Simon, éste último ha comentado últimamente que se debe considerar la condición de símbolo ampliamente de manera que los sistemas conexionistas, así como los robots de R. Brooks, se puedan incluir en los sistemas de símbolos físicos:

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).

De esta manera, la hipótesis de sistemas de símbolos físicos se continúa considerando válida. Bajo esta consideración, las redes neuronales se pueden ver como un mecanismo más de representación del conocimiento.

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, págs. 104-105).

Como se verá, en el desarrollo de la asignatura se utilizará el punto de vista de Newell y Simon. Esto es, se presentarán los elementos básicos de la inteligencia artificial centrados en los temas de búsqueda y representación del conocimiento. Además, también siguiendo a estos autores, se verán las redes neuronales como una forma de representación del conocimiento.

2.3. Sistemas emergentes

Otra alternativa a la hipótesis de Newell y Simon es la planteada por las teorías emergentes.

Se parte del hecho de que la solución a un problema no se obtiene de un único sistema individual, sino que emerge de las actividades de unos agentes independientes (y relativamente simples aunque normalmente especialistas).

Las tecnologías de agentes, que actualmente están en auge e influyen en el diseño de sistemas de computación distribuida (en especial de Internet), ofrecen modelos de cooperación social, de manera que se pueden construir sistemas complejos a partir de la cooperación de agentes individuales.

Estas ideas aparecen también en las ciencias cognitivas, donde se afirma que la mente se organiza en un conjunto de unidades funcionales especializadas.

Aplicaciones

La emergencia se aplica a diferentes niveles. Es en el ámbito de individuos donde el comportamiento de un individuo se obtiene a partir de la interacción de módulos independientes, o bien en el ámbito de poblaciones (conjunto de individuos). Este último caso es el que encontramos en los robots diseñados por el equipo de R. Brooks que hemos mencionado antes. Se han diseñado y se han construido robots simples que interaccionan como agentes autónomos solucionando problemas en un laboratorio. No existe en éstos un algoritmo central, sino que la cooperación emerge como resultado de las interacciones distribuidas y autónomas de los individuos. Aplicaciones parecidas se desarrollan en Internet. Aquí tenemos que las unidades independientes son los agentes. Cada uno con sus propias capacidades. Es el conjunto de agentes lo que define la conducta de un sistema multiagente.

2.4. Sistemas basados en las teorías de la acción situada y la inteligencia artificial corpórea

[...] 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 diferente de ver la inteligencia artificial es mediante las teorías de la acción situada².

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

Las **teorías de la acción situada** argumentan que no se ha de ver la inteligencia como un proceso de construcción y evaluación de modelos del mundo, sino como un proceso menos estructurado de actuar en el mundo y responder a los resultados obtenidos. Por lo tanto, se da más importancia a la habilidad de actuar que a la de explicar estas acciones; se prima la conducta al razonamiento.

El hecho de subrayar la importancia de las acciones en sí en detrimento de cómo se determina cuál es la que se debe realizar, ha llegado a plantear la inteligencia sin representación. Esto, evidentemente, se puede ver relacionado con los sistemas que utilizan redes neuronales donde, como ya se ha dicho, no existe representación explícita.

Los **sistemas reactivos** (también denominados *sistemas estímulo-respuesta*) son aquellos que no tienen un estado interno y que simplemente reaccionan a los estímulos existentes en el entorno.

La visión de actuar en el mundo y de responder a lo que sucede implica dar mucha importancia a nuestra habilidad para sentir (*sense*) lo que hay a nuestro alrededor y responder de manera rápida y continuada a los cambios que se suceden allí. Esto lleva a subrayar la importancia de los cuerpos y de los sentidos que nos sitúan en el entorno en lugar de darla a los procesos de un nivel más alto que razonan sobre este entorno.

En esta teoría, a diferencia del caso de la hipótesis de los sistemas de símbolos físicos, la instanciación del sistema de símbolos físicos resulta relevante. Para llevar a cabo una acción inteligente en el mundo es necesario un cuerpo físico que permita al agente estar completamente situado en el mundo. La inteligencia se obtiene a partir de la actividad práctica de captar el mundo de cada día. Es la llamada *inteligencia artificial corpórea (embodied)*.

Lectura complementaria

R. Brooks tiene un trabajo titulado precisamente así:

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

Sin embargo, no todo el mundo que critica el hecho de tener los símbolos como únicos elementos significativos del problema está de acuerdo con la necesidad de la inteligencia artificial corpórea. Hay quien afirma, como Collins, que más que hablar de inteligencia artificial corpórea, deberíamos hablar de inclusión social³. En esta interpretación, no resulta necesaria la presencia física mediante un cuerpo, sino que lo que resulta imprescindible es la relación con los agentes sociales de este entorno. Es imprescindible la socialización de los sistemas para conseguir su inteligencia.

De hecho, Collins afirma que el mundo conceptual ya está “situado”:

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.

Y más adelante en el mismo trabajo, Collins señala:

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).

Aplicaciones

La importancia de actuar frente a la de razonar ha influido en las redes neuronales y en los sistemas basados en agentes, y se encuentra, además, en el corazón de los ataques a los sistemas simbólicos. En lugar de estos sistemas se defienden principalmente los procesos de aprendizaje y adaptación.

Estos modelos han influido también en los trabajos de R. Brooks. En estos trabajos se argumenta que la inteligencia artificial se equivoca empezando por los procesos de razonamiento a niveles altos. Brooks propone empezar con la construcción y evaluación de sistemas pequeños y simples (que operan como insectos) y que permiten estudiar los procesos mediante los que tanto las criaturas sencillas como las complejas actúan en el mundo. Después de haber desarrollado sistemas de este tipo, ahora se desarrollan sistemas más complejos.

Cabe señalar que estas ideas han influido también en aspectos de la llamada *inteligencia artificial simbólica* (la que sigue la hipótesis de los sistemas de símbolos físicos). Por ejemplo, en la planificación reactiva⁴. En este tipo de planificación se abandonan las técnicas tradicionales que intentan construir un plan completo para llegar al objetivo desde un estado inicial. Estos sistemas no

⁽³⁾En inglés, *social embedding*.

Lectura complementaria

Collins se refiere al libro de Dreyfus siguiente:

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

Ved también

Los problemas de planificación se definen en el apartado 1.1.1 del módulo “Resolución y problemas de búsqueda”.

⁽⁴⁾En inglés, *reactive planning*.

funcionan adecuadamente porque suelen surgir problemas no previstos cuando el plan se ejecuta. Los sistemas de planificación reactiva se implementan con un ciclo de:

- 1) construcción de planes parciales;
- 2) actuación del plan y
- 3) reevaluación de la situación para nuevos planes.

El énfasis en la acción resulta cada vez más importante en la inteligencia artificial. De aquí el resurgimiento de la robótica y los problemas de percepción. Esto explica también el interés que ha tenido Internet: los agentes para web (o *softbots*) que realizan tareas útiles en la red Internet.

2.5. Métodos y técnicas de la inteligencia artificial

Aunque existen diferentes puntos de vista, y los objetivos del campo a largo plazo pueden resultar ambiciosos, la inteligencia artificial se dedica a la construcción de programas. Para construirlos se ha desarrollado un conjunto de métodos y técnicas que son los elementos básicos para construir sistemas más complejos.

Un esquema de los métodos y técnicas de la inteligencia artificial es éste:

- Resolución de problemas y búsqueda
- Sistemas basados en el conocimiento
- Aprendizaje
- Inteligencia artificial distribuida

Estos temas constituyen lo que se puede llamar el núcleo central de la inteligencia artificial. Estos elementos se aplican a la resolución de problemas en muchos campos diferentes (como la biología, la economía, la educación o la historia). Entre los campos a los cuales se aplican estas técnicas existen unos que, de hecho, se puede decir que forman parte de la misma inteligencia artificial porque influyen directamente en su desarrollo. Estos campos son:

- El lenguaje natural
- La visión
- La robótica
- El reconocimiento del habla

No todo el mundo está de acuerdo con esta separación a dos niveles.

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 esta asignatura se estudiarán algunos de los métodos y técnicas del primero grupo. En concreto, nos centraremos en los temas de resolución de problemas y búsqueda, y en los sistemas basados en el conocimiento. También se realizará una introducción a las diferentes formas de supervisión de los algoritmos de aprendizaje computacional: supervisado, no supervisado y por refuerzo.

En la asignatura *Aprendizaje computacional* es donde se estudiarán con más detalle los diferentes algoritmos de aprendizaje y también se introducirá la inteligencia artificial distribuida a partir de los sistemas multiagentes.

Entre las aplicaciones que existen actualmente en torno a las técnicas de aprendizaje podemos subrayar la prospección de datos⁵, también conocida por descubrimiento de conocimiento de bases de datos⁶. Esto corresponde a extraer información potencialmente útil, que antes no se conocía, pero que se encontraba implícita en unos datos.

⁽⁵⁾En inglés, *data mining*.

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

La inteligencia artificial distribuida corresponde a situaciones en las que existen diferentes sistemas que interactúan para resolver un problema común. Dentro del campo hay dos áreas principales: la resolución de problemas de manera distribuida⁷ y los sistemas multiagentes⁸. En la primera área se estudia cómo dividir una tarea en módulos que cooperan y comparten información sobre un problema y sobre las soluciones. Sin embargo, en este caso toda forma de interacción entre módulos (las estrategias de cooperación y coordinación) se encuentra prefijada e incorporada en el sistema. En la segunda área se estudia la construcción y la conducta de una colección de agentes autónomos que intentan resolver un problema. Estos sistemas pueden ser heterogéneos, esto es, con capacidades de razonamiento diferentes.

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

⁽⁸⁾En inglés, *multiagent systems*.

3. Algunas aplicaciones

Hasta ahora se han desarrollado muchos sistemas que utilizan las técnicas desarrolladas en el campo de la inteligencia artificial. Estos sistemas han sido construidos para ser utilizados en entornos muy diferentes (de entornos industriales a objetos de uso cotidiano, desde aplicaciones médicas a herramientas para ayudar al aprendizaje, por ejemplo, de materias relacionadas con la historia). En este apartado mencionaremos algunas aplicaciones. Primero, repasaremos algunos de los sistemas que fueron innovadores en el momento de su construcción y acabaremos con algunos sistemas más recientes (entre paréntesis se incluirán los investigadores principales de cada sistema y el año de publicación del resultado):

1) Logic Theorist (Newell, Shaw, Simon, 1957). Era un demostrador automático de teoremas. Esto es, demostraba un teorema a partir de un conjunto de axiomas dado. Entre lo que demostró encontramos teoremas del segundo capítulo del libro de Russell y Whitehead titulado *Principia Mathematica*. Además, de uno de estos teoremas realizó una demostración más corta que la dada en el libro. Como ya hemos señalado antes, el Logic Theorist fue presentado en el encuentro de Dartmouth del verano de 1956.

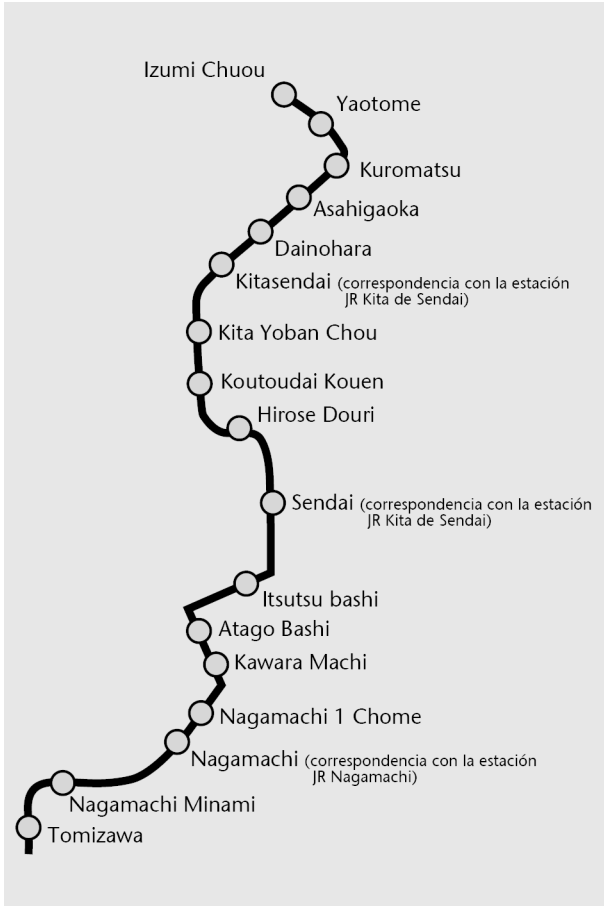
2) Mycin (Shortliffe y Buchanan, 1975). Era un sistema experto diseñado para ayudar en la recomendación de terapias apropiadas para pacientes con infecciones bacterianas. El sistema utilizaba reglas para razonar que, con el fin de modelizar la incertidumbre, incorporaban los llamados *factores de certeza*.

3) AM (Lenat, 1977). Era un programa experimental para construir teorías de manera automática en el campo de las matemáticas. A partir de un conjunto de conceptos elementales de la teoría de conjuntos y de la aritmética (y de la definición de estos conceptos), AM producía nuevos conceptos y conjeturas a partir de unos mecanismos generadores (operadores). Además, mediante un conjunto de reglas seleccionaba de entre los conceptos generados aquellos que creía que podían resultar interesantes. Entre los conceptos que generó, se encuentra el de número primo. Constituía uno de los primeros sistemas de descubrimiento automático.

4) Prospector (Duda, 1979). Era un sistema experto construido para ayudar en la localización de depósitos de minerales como el cobre y el uranio. Fue uno de los sistemas que contribuyó al renombre de la inteligencia artificial hacia los años ochenta al recomendar la excavación en un lugar donde se encontró un depósito de molibdeno muy grande.

5) Metro de Sendai (Yasunobu, 1987). El sistema de control para la conducción automática del metro de la ciudad japonesa de Sendai fue desarrollado por Seiji Yasunobu del laboratorio de desarrollo de sistemas de Hitachi. El sistema está construido mediante reglas difusas. Actualmente, existen sistemas basados en reglas difusas en muchos aparatos de uso cotidiano (lavadoras, cámaras de fotografiar, aparatos de aire acondicionado, etc.).

Figura 1



Plano del metro de la ciudad de Sendai

6) EQP (McCune, 1997). El sistema EQP demostró en 1996 el llamado *problema de Robbins*. Este problema, que fue planteado por H. Robbins en los años treinta, es muy conocido en el álgebra booleana y hasta ahora era un problema abierto (que no se había podido resolver). El sistema EQP es un programa de demostración automática para lógica ecuacional. La solución fue encontrada de manera automática entre septiembre y octubre de 1996.

7) Deep Blue (IBM, 1997). Por primera vez un ordenador jugador de ajedrez ganó al campeón humano mundial, G. Kasparov, en Nueva York durante el mes de mayo del año 1997. El programa utiliza algoritmos de búsqueda paralela basados en los propuestos por C. Shannon, además de hardware dedicado. Deep Blue posee conocimiento del dominio incorporado en la función que utiliza para evaluar los diferentes tableros que plantean las distintas opciones (función construida por M. Campbell con la colaboración de los jugadores de

Lectura complementaria

Encontraréis más información sobre el sistema de control del metro de Sendai y sobre Deep Blue, respectivamente, en las direcciones siguientes:

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

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

ajedrez J. Benjamin y M. Illescas). Además, Deep Blue incorpora bases de datos de finales del juego que permiten jugar un juego perfecto cuando está en aquellas posiciones.

8) DARPA Grand Challenge (Stanford University, 2005). El DARPA Grand Challenge consiste en una carrera de vehículos autónomos que tienen que llegar de un punto de Estados Unidos a otro punto sin intervención humana y disponiendo únicamente de un listado de puntos intermedios entre el principio y el final del circuito. En 2005, por primera vez, cinco vehículos autónomos consiguieron completar los 212 kilómetros de que consta el circuito, con un vehículo de la Stanford University como ganador, con un tiempo de 6 horas y 53 minutos.

9) DARPA Urban Challenge (CMU, 2007). A diferencia del DARPA Grand Challenge de 2005, esta prueba se desarrollaba en una área urbana, en un circuito de 96 kilómetros, donde los vehículos tenían que cumplir la normativa de tránsito, además de evitar los choques con los otros vehículos de la carrera, tanto vehículos autónomos como conductores profesionales. En las ediciones anteriores, los vehículos no encontraban a otros durante el recorrido, a pesar de que las condiciones físicas del terreno eran más desafiantes. El vehículo autónomo ganador fue diseñado por la Carnegie Mellon University (CMU) y completó el circuito en un tiempo de 4 horas 10 minutos.

10) Watson (IBM, 2011). En 2011, un sistema informático diseñado por IBM, capaz de responder a preguntas formuladas en lenguaje natural, ganó a los dos mejores participantes humanos de un concurso televisivo llamado Jeopardy! Se trata de un concurso con preguntas de varias temáticas, como, por ejemplo, historia, idiomas, literatura, cultura popular, bellas artes, ciencias, geografía y deportes. El programa informático tenía acceso a doscientos millones de páginas de contenido estructurado y no estructurado, pero no disponía de conexión a internet durante el juego. El proyecto fue desarrollado por un equipo de investigación con la tecnología DeepQA⁹ de IBM.

⁽⁹⁾En inglés, QA hace referencia a *Question Answering* (respuesta de preguntas).

11) AlexNet (Krizhevsky, 2012). El diseño e implementación mediante GPU de una red neuronal convolucional, formada por ocho capas (cinco convolucionales y tres totalmente conectadas), supuso una mejora muy significativa en una competición de reconocimiento visual de imágenes llamada ImageNet: se pasó de un error del 26,1% a un error del 15,3%. La tarea consiste en designar las cinco categorías más probables de una imagen entre las mil posibles categorías definidas. A partir de 2015, con redes cada vez más profundas, se consiguió un error inferior a la tasa de error humana, estimada en un 5%.

12) AlphaGo (Google DeepMind, 2015). En octubre de 2015, un programa informático derrotaba por primera vez a un jugador profesional de Go, un juego de estrategia de tablero para dos jugadores muy popular en Asia oriental, que se practica en un tablero de 19x19 y cuya complejidad es significativamente superior a la del ajedrez. El programa utiliza un algoritmo de búsqueda en árbol para hallar los movimientos, basándose en el conocimiento previamente aprendido mediante una red neuronal a partir de las partidas jugadas por los humanos y por el mismo ordenador.

Lectura complementaria

AlphaGo tuvo un eco mediático importante, con la producción incluso de un documental que recibió varios premios en festivales de cine.
[https://
www.alphagomovie.com](https://www.alphagomovie.com)

4. Algunas características de los sistemas inteligentes

Hasta ahora se han construido muchos sistemas inteligentes que se aplican en entornos muy diferentes. De todos modos, a pesar de la diversidad de sistemas y de campos de aplicaciones, existen algunos elementos comunes en todos estos programas. A continuación se describen estos elementos. Resulta importante subrayar que no todos los sistemas los aplican todos, y que la importancia que se da a cada elemento depende mucho del sistema.

Las características de estos sistemas son las siguientes:

- 1) **Usan información simbólica.** La información que han de tratar los sistemas es sobre todo simbólica. Deben razonar sobre cosas que ocurren (hechos), conceptos abstractos y han de sacar conclusiones.
- 2) **Utilizan descripciones del dominio de la aplicación.** Para poder encontrar una solución, los sistemas deben conocer el entorno del problema. Han de saber qué hay, cómo pueden actuar, cómo evoluciona el entorno.
- 3) **Utilizan datos incompletos, inexactos o con conflicto.** La información que han de tratar los sistemas no se encuentra en las condiciones ideales que se querría. A menudo se debe tratar el hecho de que falta información, o de que los datos recibidos se encuentran en conflicto (entre éstos o con el conocimiento que el sistema ya tiene incorporado). También se encuentran problemas de incertidumbre (valores aproximados, conceptos difusos, etc.).
- 4) **Utilizan métodos heurísticos.** Los programas realizan lo que se les dice, pero en el caso de muchas aplicaciones de inteligencia artificial lo que tenemos implementado son indicaciones de cómo se debe encontrar la solución, más que descripciones detalladas con todos los pasos que se deben seguir. Los métodos heurísticos nos dan una idea de cómo hemos de buscar la solución, pero esto no nos garantiza el éxito. A menudo no existe la mejor solución, pero se encuentra una suficientemente buena para operar.
- 5) **Son adaptativos.** Cuando el entorno en el que se encuentra un sistema cambia, el sistema también debe cambiar a fin de que su comportamiento continúe siendo el adecuado.

Ved también

Podéis ver algunas aplicaciones de la inteligencia artificial en el apartado 3 de este módulo.

Por ejemplo...

... muchos de los sistemas basados en el conocimiento (podéis ver el módulo "Sistemas basados en el conocimiento" de esta asignatura) no son adaptativos; algunos sistemas reactivos no utilizan información simbólica.

Ejemplo de descripción del dominio de aplicación

En un sistema de diagnóstico para un determinado tipo de aparatos, se deberán conocer las relaciones entre los síntomas y las causas, además habrá que saber cómo se comportan los componentes del aparato cuando todo funciona correctamente.

Actividades

1. Podéis leer el artículo de A. Turing publicado en el año 1950 en la revista *Mind* (núm. 59, págs. 433-460) y considerad las predicciones que realizó para el año 2000.
2. Considerad las diferentes maneras de entender la inteligencia artificial. ¿Cuál os parece más acertada?
3. ¿Creéis que es cierto, tal como afirma Simon, que la inteligencia artificial es una ciencia experimental?
4. ¿Cuál de las cuatro definiciones alternativas sobre la IA, considerando conducta frente a razonamiento y racionalidad frente a comparación con el comportamiento de las personas, os parece más adecuada?
5. Considerar los problemas que se presentan a continuación. El tres en raya y el parchís son programas de juegos, y la manera como tratarlos se describe en el módulo “Resolución de problemas y búsqueda” de esta asignatura. Lo mismo ocurre con el problema de la integración de expresiones. El caso de los circuitos es considerado como un ejemplo del razonamiento basado en modelos. En el módulo “Sistemas basados en el conocimiento” de esta asignatura se dedica un apartado a este tema. De todos modos, antes de pasar a los módulos podéis responder a las preguntas siguientes.
 - a) En el juego del tres en raya, ¿cómo lo haríais para elegir la mejor jugada? ¿Y en el caso del parchís?
 - b) Considerad el problema de la integración de expresiones (por ejemplo, cuál es la integral de $x^2 + 3e^x$). ¿Cómo resolveríais este problema?
 - c) Disponéis de una fábrica de construcción de circuitos y queréis construir un sistema que compruebe el funcionamiento de cada circuito y, cuando no funcione, determinar qué es lo que falla. ¿Cómo trataríais este problema?

Bibliografía

Bibliografía básica

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

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

Bibliografía complementaria

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

Collins, H. M. (1996). . "Embedded or embodied? A review of Hubert Dreyfus' 'What Computers Still Can't Do'". *Artificial Intelligence* (núm. 80, págs. 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, págs. 119-128).

Newell, A.; Simon, H. A. (1976). "Computer Science as Empirical Inquiry: Symbols and Search". *Communications of the ACM* (vol. 3, núm. 19, págs. 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, págs. 95-127).