

LA INTELIGENCIA ARTIFICIAL EN NUESTROS DIAS

Lorenzo Moreno Ruiz
Dpto. Física Fundamental y Experimental
Universidad de la Laguna

*... en el fuego de lo que fue
arde lo que será*

Louis Aragon

INTRODUCCION

El hombre, a lo largo de su presencia en la tierra, ha tenido que desarrollar su ingenio para poder subsistir en un medio hostil. Ha tratado de crear artilugios o herramientas que fueran prolongaciones de sus miembros, y le posibilitaran controlar de forma creciente ese medio agresivo, permitiéndole dedicarse a otras tareas más trascendentes o creativas que no estuvieran tan directamente relacionadas con la elemental subsistencia.

Ese afán del hombre por crear cada vez medios más poderosos para imitar o controlar la Naturaleza, además de su deseo irrefrenable por saber, conocer y dar respuesta a las preguntas que se ha ido planteando a lo largo de los siglos, junto con otros deseos, más prosaicos y menos confesables, guiados en algunos casos por una ambición desmedida o por soberbia, le han permitido ir perfec-



cionando los autómatas que ha ido diseñando y construyendo a lo largo de su historia. Podemos citar, a modo de resumen, desde los autómatas voladores contruidos en la antigua China, tradición hermética en el Renacimiento, el hombre máquina de La Mettrie (que afirmaba que el cuerpo humano no era más que un reloj), el regulador centrífugo de Watt en la máquina de vapor (precursora de la revolución industrial), la máquina de Babbage, los autómatas de Torres Quevedo, el computador digital, el robot... hasta la declaración de Marvin Minsky hace algunos años de que " el cerebro resulta ser una máquina de carne", o la de los partidarios de la Inteligencia Artificial fuerte de que "la mente es al cerebro lo que el programa al ordenador".

Pamela McCorduck en su libro "Máquinas que piensan" afirma:

"La Inteligencia Artificial (IA) es la más reciente manifestación de un impulso permanente del hombre por crear artefactos que imiten nuestra propiedad esencial, la inteligencia".

Actualmente los científicos, psicólogos, filósofos e ingenieros están mostrando un interés creciente hacia lo que John McCarthy acuñó en 1956 como IA, con ocasión de una reunión histórica en el Dartmouth College¹. Es ésta una disciplina joven y que desde su aparición ha motivado multitud de polémicas, internamente entre sus diferentes escuelas o corrientes y, sobre todo, entre sus defensores y sus más acérrimos enemigos. Una disciplina de la que todo el mundo se atreve a opinar sin haberse informado o meditado lo suficiente, ocurriendo lo mismo que en otras áreas del conocimiento como la Política o la Informática.

En las próximas líneas trataré de presentar aquellos planteamientos realizados en torno a la IA que me han parecido más interesantes, tratando de exponer las razones por las que los filósofos, psicólogos, matemáticos, biólogos... se han acercado a ella; esto permitirá que al final del apartado pueda presentar las definiciones de IA que me han parecido más adecuadas, sin exclusiones interesadas.

En el siguiente apartado, abordaremos de forma sucinta los temas en los que se haya implicada la IA.

Finalmente, en el último apartado expondremos una historia de la IA, con las principales vicisitudes por las que ha ido pasando esta apasionante disciplina.

¹ Algunos de los asistentes que se reunieron en el verano del 56 en el Dartmouth College de Hanover, New Hampshire eran: John McCarthy, profesor ayudante en Dartmouth, Marvin Minsky profesor de matemáticas y neurología en Harvard, Claude Shannon matemático de los Laboratorios Bell Telephone y Nathaniel Rochester, director de investigación de información en el Centro IBM en Poughkeepsie en New York, Allen Newell de la RAND Corporation de Santa Mónica y Herbert Simon del Instituto Carnegie de Pittsburgh.



DEFINICIÓN DE IA

G. N. Reeke y G. M. Edelman, profesores de la Universidad Rockefeller, en su artículo "Cerebros Reales e Inteligencia Artificial" afirman que:

"La IA es uno de esos términos que tienen un significado tan evidente que rara vez se define con cuidado. Como resultado, ha llegado a tener una interpretación tan amplia que agrega confusión al debate sobre sus méritos y perspectivas. Algunos han llegado tan lejos como para definir la IA como cualquier cosa en la que trabaja la gente de IA".

Evidentemente una definición recurrente como la anterior nos aclara bien poco.

Continúan estos autores indicando que:

"Para ayudarnos a ver más claramente estas cuestiones, excluirémos las empresas que poseen fines de ingeniería pura, es decir, los proyectos que intentan desarrollar soluciones computacionales efectivas a problemas que ya están bien comprendidos en principio".

Se refieren a eliminar del término IA las aportaciones realizadas en el campo de las matemáticas y la ingeniería, como la programación lógica y los sistemas expertos. Sin embargo, si algún éxito cabe apuntarle a la IA en el presente es precisamente la aplicación de los sistemas expertos en una gran variedad de campos de aplicación como la medicina, la geología, la economía,... o la programación lógica en la demostración de teoremas. Trataremos, pues, de encontrar otras definiciones que nos aclaren alguna cosa y no traten de excluir ninguna faceta de la IA.

Robert Sokolowski, profesor de filosofía en la Universidad Católica en EEUU, en su artículo "Inteligencia Natural e Inteligencia Artificial", se hace una pregunta que es clave para la definición de IA:

"¿En qué sentido usamos la palabra artificial cuando hablamos de IA? Los críticos, los que calumnian la idea, la identifican con aquello que parece ser pero que realmente no es lo que parece. Por el contrario, los que apoyan la idea de la IA dirán que el término se refiere a aquello que es fabricado pero que después se comporta como lo que parece ser, es decir, dirán que las máquinas que piensan son artefactos obviamente ejecutadas por seres humanos, pero que una vez corriendo, las máquinas piensan. Su pensamiento puede ser diferente al de los seres



humanos, de la misma manera que un avión es diferente a un pájaro aunque ambos vuelan.

¿Estamos obligados a demostrar que las máquinas piensan realmente? Quizá no, nadie ha probado que los aviones realmente vuelan. Solamente vemos que lo hacen. Si las máquinas que piensan exhiben la actividad de pensar ¿Por qué no debemos admitir que son realmente inteligentes? El problema es que el pensamiento no es tan visible y palpable como lo son el movimiento y el vuelo".

Alan Turing propuso en 1950 un test para la detección de inteligencia en una máquina. El denominado test de Turing es una prueba que un interrogador humano realiza, vía un teletipo, para determinar si se trata de un computador o una persona quien está respondiendo.

El test de Turing evita deliberadamente la interacción física entre el interrogador y el computador, dado que sería irrelevante la presencia física para la detección de inteligencia. Sin embargo, el denominado test de Turing total añade además la capacidad de percepción visual y las habilidades manuales.

Hay que indicar que cuando Turing formuló su famoso test en el artículo "Computing Machinery & Intelligence", los computadores no habían sido inventados aún, sólo se conocía el ingenio Analítico de Babbage (1830), aunque él conocía los trabajos de Norbert Wiener sobre Cibernética, además de los trabajos de John Von Neumann sobre los Computadores Digitales.

La Profesora Putnam de la Universidad de Harvard en su artículo "Mucho ruido por muy poco" nos presenta la siguiente definición sobre IA:

"El diseño de Computadores es una rama de la Ingeniería y la IA es una subrama de esta Ingeniería, cuyo objetivo es desarrollar software que permita a los computadores simular o duplicar los logros de los que intuitivamente reconocemos como inteligencia.

Estamos tentados de decir que en principio puede ser posible modelizar la mente o el cerebro con una computadora digital y el software apropiado.

Tratar de programar en la máquina toda la información que posee un humano requeriría varias generaciones de investigadores para formalizar la información y no está claro que el resultado vaya a ser algo más que un gigantesco sistema experto. Una inteligencia de este tipo sería con toda probabilidad muy poco imaginativa".

En contraposición, B. Mazlish profesor de la Universidad de Yale en su libro "La cuarta discontinuidad" nos plantea la siguiente hipótesis:



"Los hombres pueden o no ser máquinas, pero lo que sí es una parte de la realidad es que los humanos han ido creando una civilización cada vez más mecanizada, en especial desde la Revolución industrial y tal vez aspiran a ser máquinas".

El profesor Mazlish mantiene que las tres ideas: Copernicana (del lugar no especial del hombre en el universo) con el desarrollo del método científico, la Darwiniana del producto, aparentemente final, de la evolución del reino animal y la Freudiana de la relevancia del subconsciente en los procesos mentales, han supuesto una serie de puentes que han salvado las siguientes discontinuidades: hombre - cosmos, hombre - animal, racional - irracional. Pues bien, él plantea que la aparición de los computadores ha permitido crear un cuarto puente hombre - máquina en el sentido del fisiólogo británico Thomas Henry Huxley (1825-1895), para el que convertirse en una máquina era llegar a la perfección moral, dado que éstas son incapaces de cometer errores de funcionamiento.

Mazlish plantea la siguiente pregunta:

"¿En qué máquina quiere convertirse Huxley? Una de las líneas actuales sobre IA dice que se puede conseguir que las máquinas piensen como los humanos, es decir, no habrá ninguna diferencia entre el hombre y el ordenador porque ambos serán pensantes. Entonces, puesto que los hombres cometen errores, las máquinas deberían cometer errores (paradoja que hace que la perfección tan deseada por Huxley nunca pueda ser alcanzada convirtiéndose en máquina).

Otra posibilidad es que las máquinas son o pueden llegar a ser más racionales y más perfectas que nosotros sin llegar a ser totalmente perfectas. Entonces, podríamos acercarnos asintóticamente al deseo de Huxley, aunque éste no podría ser alcanzado con toda perfección".

John R. Searle, profesor de filosofía en la Universidad de California, es, sin duda, uno de los máximos detractores de la IA. En su artículo: ¿Es la mente un programa informático? se hace la siguiente pregunta:

"¿Podría una máquina pensar justamente en virtud de que ejecuta un programa informático?"

El profesor Searle califica como IA fuerte a los investigadores que creen que diseñando programas adecuados y dotándolos de las oportunas entradas y salidas están literalmente creando mentes. Por contra, califica como IA débil a los investigadores que adoptan un punto de vista más cauteloso, es decir, que los



modelos informáticos resultan útiles para estudiar la mente, de igual manera que lo son para el estudio de la meteorología o la biología.

Los partidarios de la IA fuerte, en palabras de R. Penrose en "La nueva mente del Emperador", mantienen que:

"La actividad mental consiste simplemente en llevar a cabo alguna secuencia bien definida de operaciones, frecuentemente llamada algoritmo. Todas las cualidades mentales - pensamiento, sentimiento, inteligencia, comprensión, consciencia - son características del algoritmo a ejecutar por el cerebro".

Deberíamos aquí presentar con rigor el concepto de algoritmo. Alan Turing demostró que cierta clase de problemas matemáticos no pueden resolverse mediante ningún procedimiento fijo y definido. Por procedimiento definido, entendía algo que podría hacerse mediante la máquina automática abstracta que proponía, y que ha pasado a conocerse como máquina de Turing.

El término algoritmo intervenía en matemáticas solamente con relación a algoritmos concretos, pero no existió una definición rigurosa de algoritmo hasta 1937, en que Turing publicó su artículo en los Proceedings of the London Mathematical Society.

La máquina de Turing constaba de un control finito, una cinta de entrada dividida en celdas, una cabeza lectora que muestreaba una celda de la cinta en cada instante. La cinta es finita hacia la izquierda, sin embargo es indefinida a la derecha. Cada celda de la cinta puede mantener exactamente un símbolo de un número finito de ellos. En un movimiento de la máquina de Turing, dependiendo del símbolo muestreado y del estado en que se encuentre el control, la máquina cambia de estado, imprime un símbolo no blanco sobre la celda y mueve la cabeza un lugar hacia la derecha o la izquierda.

Una máquina reconocerá una cadena de símbolos, escrita inicialmente sobre la cinta, si a través de los movimientos especificados en la máquina, dicha cadena es reconocida, y esto se consigue si la máquina alcanza, tras el movimiento, un estado final establecido.

Pues bien, todos los algoritmos pueden ejecutarse sobre una máquina de Turing. Esta hipótesis puede compararse con la ley de la conservación de la energía en el sentido de que su validez está basada en la experiencia.

Llamaremos máquina universal de Turing a una computadora de propósito general que permite simular cualquier máquina de Turing, incluyendo ella misma. Esta máquina contendrá dos cintas. La nueva cinta almacena la codificación de cualquier máquina de Turing.

Diremos que una función f es calculable si $f(x)$ puede evaluarse para cualquier x en un número finito de pasos mediante una máquina de Turing.



Hay funciones que no son calculables, como ejemplo típico podemos citar el problema de la parada de las máquinas de Turing. Sea (T,x) una máquina de Turing T con su entrada x . Diremos que (T,x) para si T con la entrada x para después de un número finito de pasos. Sea $f_h(T,x)$ una función definida para todas las máquinas de Turing y todas las configuraciones de entrada

$$\begin{aligned} f_h(T,x) &= 1 \text{ si } (T,x) \text{ para} \\ f_h(T,x) &= 0 \text{ si } (T,x) \text{ nunca para} \end{aligned}$$

Esta función así definida no es calculable. Esto implica que el problema de determinar si (T,x) se detiene, cuando (T,x) es cualquier máquina de Turing con cualquier configuración de entrada es irresoluble.

Una vez definido el concepto de algoritmo, prosigamos con la pregunta realizada por el profesor Searle.

Mantiene que la simple ejecución de un algoritmo no implica que haya tenido lugar ninguna comprensión. Para corroborar su afirmación, propuso el famoso ejemplo de la habitación china consistente en lo siguiente:

"Tomemos un idioma que no comprendemos, por ejemplo el chino. Supongamos que nos instalan en una habitación que contiene cestas repletas de símbolos chinos. Supongamos también que nos proporcionan un libro de instrucciones en español con reglas que estipulan como han de emparejarse unos símbolos chinos con otros. Las reglas permiten reconocer los símbolos puramente por su forma y no requieren que comprendamos ninguno de ellos. Las reglas serían del tipo: Tome un signo de la cesta número 1 y colóquelo al lado de un símbolo de la cesta número 2. Imaginemos que personas situadas fuera de la habitación y que sí comprenden el chino nos van entregando pequeños grupos de símbolos, y que, en respuesta, nosotros manipulamos los símbolos de acuerdo con las reglas del libro y les entregamos pequeños grupos de símbolos. Ahora, el libro de instrucciones es el programa informático, las personas que lo escribieron son los programadores y nosotros somos el ordenador. Las cestas llenas de símbolos constituyen la base de datos, los pequeños grupos que nos son entregados son preguntas, y los grupos que nosotros entregamos son las respuestas. Supongamos ahora que el libro de instrucciones esté escrito de modo tal que nuestras respuestas a las preguntas resulten indistinguibles de las de un chino nativo. Estamos superando el test de Turing en lo que se refiere a la comprensión del chino, y al mismo tiempo ignoramos totalmente el chino".



Como conclusión, Searle mantiene que:

"La mera disponibilidad de la sintaxis no basta para disponer de la semántica. La mera manipulación de símbolos no es suficiente para garantizar el conocimiento de sus significados".

Sin embargo, a través de la manipulación de símbolos se han resuelto una gran cantidad de problemas complejos de juegos de damas, de ajedrez, de demostración de teoremas, y a medida que las posibilidades tecnológicas nos han ofrecido procesadores más rápidos y más potentes, los programas se han ido perfeccionando y cada vez son mejores los resultados.

R. Penrose en su libro "La nueva mente del Emperador" afirma que:

"...incluso la ejecución de un programa de ordenador bastante sencillo sería normalmente algo extraordinariamente largo y tedioso si fuera realizado por seres humanos manipulando símbolos. La dificultad es mayor con un supuesto programa de ordenador que se supone tiene suficiente complicación para igualar a un cerebro humano y, por lo tanto, superar perfectamente el test de Turing. Podemos imaginar que la operación de este programa para dar respuesta a incluso alguna pregunta bastante simple del test de Turing, podría suponer tantos pasos que no habría posibilidad de que ningún ser humano llevara a cabo el algoritmo a mano en el plazo de tiempo de una vida humana. ...no me resulta inconcebible que pudiera haber algún grado crítico de complicación que fuera necesario alcanzar en un algoritmo para que éste muestre cualidades mentales. Quizá este valor "crítico" es tan grande que ningún algoritmo pudiera concebiblemente ser ejecutado a mano por ningún ser humano a la manera imaginada por Searle".

El propio Searle ha replicado a esta última objeción modificando su habitación china y ampliándola para incluir a la India, con toda su población ocupada en la manipulación de símbolos, y argumentando que los resultados serían los mismos.

Y entonces Penrose afirma:

"...¿No son todos estos indios individuales más semejantes a las neuronas individuales del cerebro de una persona que al cerebro global? Nadie sugeriría que las neuronas, cuyas descargas constituyen aparentemente la actividad física de un cerebro en el acto de pensar, comprendan ellas mismas individualmente lo que la persona está pensando".



do, así que ¿ Por qué esperar que los indios individuales comprendan las historias chinas?"

Hemos llegado a los fundamentos de la vertiente conexionista de la IA, en la que están interesados entre otros los biólogos, los psicólogos y los neurofisiólogos.

Hubert L. Dreyfus y Stuart E. Dreyfus, profesores de la Universidad de California en Berkeley, afirman en su artículo "Fabricar una mente versus modelar el cerebro: la inteligencia artificial se divide de nuevo" lo siguiente:

"Si el conexionismo logra cumplir su promesa los investigadores tendrán que renunciar a la convicción de Descartes de que la única forma de producir conducta inteligente es reflejando el mundo mediante una teoría formal de la mente. Peor aún, se tendrá que renunciar a la intuición más básica que está en la fuente de la filosofía de que debe haber una teoría para cada aspecto de la realidad. Las redes neuronales pueden demostrar que nos comportamos inteligentemente en el mundo sin tener una teoría de ese mundo".

Es conocido que todo el procesamiento de la información en un individuo se realiza en el cerebro a través de células denominadas neuronas. La actividad computacional de cada neurona involucra una gran cantidad de mecanismos, todavía no muy bien conocidos. Se sabe que una neurona transmite información a las neuronas vecinas a través de uniones neuronales llamadas sinapsis. Una sola neurona puede tener una gran cantidad de entradas sinápticas, recibiendo de cada una de ellas señales electroquímicas que disparan una amplia variedad de reacciones. Una es la modulación de la conductividad iónica de la membrana de la neurona afectada que bien eleva el voltaje de una porción de su interior, o bien, lo disminuye. Si el voltaje combinado resultante excede el umbral de la reacción, la neurona genera una señal eléctrica. Pues bien, el cerebro humano es una red neuronal que comprende alrededor de diez mil millones de neuronas interconectadas, donde el acto de pensar no se realiza en cada neurona particular sino que está implícito en la estructura de la red neuronal.

Las premisas que caracterizan a la escuela conocida como Conexionismo en palabras de Jacob T. Schwartz, Profesor del Instituto Courant de Ciencias Matemáticas de la Universidad de Nueva York, (El nuevo conexionismo: desarrollando relaciones entre la neurociencia y la inteligencia artificial) son:

"1.- El cerebro no opera como una computadora serie convencional, sino en forma masivamente paralela de cientos de miles o millones de neuronas.



2.- Dentro del cerebro, el conocimiento no se almacena en ninguna forma que se parezca a un programa de computadora convencional sino que se almacena estructuralmente, en forma de patrones distribuidos de pesos sinápticos excitatorios e inhibitorios, cuyas magnitudes relativas determinan el flujo de las respuestas neuronales que constituyen la percepción y el pensamiento".

Dice Pamela McCorduck en su libro "Las Máquinas que piensan":

"En el siglo XX se han llevado a cabo numerosos estudios sobre la mente, tanto por parte de fisiólogos como de psicólogos, aunque los filósofos continúan preguntándose angustiosamente por la naturaleza de la mente sin prestar demasiada atención a los resultados obtenidos por los científicos. Este hecho de implicar a varias disciplinas no es necesariamente algo malo. Yo creo que la mente humana es lo suficientemente compleja como para merecer una diversidad de enfoques".

Uno de los precursores más influyentes de la moderna IA fue el neurofisiólogo Warren McCulloch. En 1943 publicó un artículo con un joven matemático llamado Walter Pitts, en el que aplicaron lógica simbólica al problema de describir lo que las redes neuronales pueden hacer. Las redes neuronales formales o redes de McCulloch y Pitts son representaciones extremadamente simplificadas de las redes neuronales naturales, que podían implementar funciones lógicas.

Estos modelos iniciales han ido mejorando a lo largo de los años, y en la década de los años 80 han hecho un progreso realmente espectacular, que ha permitido aplicar las redes neuronales multicapa u otros tipos a multitud de problemas en economía, control, robótica...

Una neurona formal es un elemento de cálculo que produce una salida que es la suma de las entradas multiplicadas por unos parámetros llamados pesos. La salida a su vez se alimenta a un elemento no lineal del tipo todo-nada o sigmoide.

Inicialmente los pesos de todas las entradas o sinapsis de cada una de las neuronas en la red se ajustan a valores arbitrarios, de manera que la respuesta inicial de la red a unos estímulos en la entrada es aleatoria. Para obtener la respuesta deseada de la red se requieren ajustar todos esos pesos a unos valores que permitan que la red responda a los estímulos conocidos con la respuesta correcta. Este proceso, que se conoce como entrenamiento o aprendizaje de la red puede ser del tipo supervisado –donde a través de un procedimiento hacia atrás las redes del tipo perceptrón multicapa obtienen los valores adecuados para los pesos minimizando una función de error al ir presentando sucesivamente a la red las entradas y salidas conocidas–, o del tipo no supervisado –donde no se le pro-



porciona ninguna realimentación para sus salidas, es decir, dado un conjunto de datos de entrada a la red se le permite jugar con ellos para intentar descubrir regularidades y relaciones entre las entradas, como por ejemplo el algoritmo de aprendizaje competitivo de Kohonen.

En estas páginas previas, hemos puesto de manifiesto la relación existente entre la IA y la ingeniería, la filosofía, la biología y la neurofisiología. Finalmente, veamos su relación con la psicología.

A finales de 1950, Allen Newell y Herbert Simon, profesores de la Universidad Carnegie-Mellon, construyeron un programa de computadora llamado GPS (General Problem Solver). Preguntas como: ¿Por qué el GPS hace una cosa u otra? hicieron plantearse a los psicólogos preguntas tales como: ¿Por qué no se habrían de usar referencias a reglas para contestar preguntas sobre lo que la gente hace cuando se enfrenta con problemas similares?

Sherry Turkle, profesora del MIT, en su artículo "Inteligencia Artificial y Psicoanálisis: una nueva alianza" manifiesta lo siguiente:

"Hacia finales de la década de 1960, la hegemonía conductista -se podía estudiar la conducta, pero no los estados interiores- se había quebrado. Había muchos factores que incidían en esta revolución científica, y el más importante era la computadora. Fue la existencia de ésta lo que proporcionó la legitimación a una forma radicalmente distinta de ver la mente (Psicología cognitiva). Si las "mentes" de las nuevas máquinas tenían estados internos, seguramente la gente también".

Evidentemente el cerebro es un órgano tan complejo que es impensable que se pueda modelar con una única red neuronal, sobre todo porque las neuronas formales que hemos utilizado en el modelo corresponden a un esquema muy simplificado de una neurona natural y además hemos supuesto que todas las neuronas son del mismo tipo, lo cual no es cierto. Sin embargo, se han conseguido éxitos importantes con este tipo de redes neuronales para modelizar problemas concretos en el cerebro, como es el caso de la comprobación por Geoffrey E. Hinton y sus colaboradores de la dislexia adquirida. El profesor Hinton de la Universidad de Toronto ha propuesto un modelo basado en una red neuronal en el que al deteriorarse algunas de sus conexiones su comportamiento es similar a lo que ocurre en varios tipos de dislexias.

Hasta aquí, hemos presentado los límites en los que se enmarca actualmente la IA, sus relaciones con otras disciplinas del saber, así como el posicionamiento de muy cualificados profesores ante esta disciplina tan controvertida.

Para finalizar este apartado, expondremos algunas de las definiciones que se han dado sobre la IA.



Según el diccionario Oxford de la mente

"La IA es la ciencia que trata de que las máquinas hagan la clase de cosas que hace la mente humana".

Una definición un poco más precisa es la dada por Elaine Rich y Kevin Knight, autores del excelente texto "Inteligencia Artificial", y dice:

"La IA estudia cómo lograr que las máquinas realicen tareas que, por el momento, son realizadas mejor por los seres humanos".

Por último, Patrick Winston, autor del libro "Inteligencia Artificial", define la IA:

"como el estudio de las ideas que permiten a las computadoras ser inteligentes"

definición en la que no aparecen los seres humanos a diferencia de las otras dos, no identificando, por tanto, la capacidad de disponer de inteligencia exclusivamente con ellos.

TEMAS ABORDADOS POR LA IA

La mayoría de los primeros trabajos en la IA hicieron gran hincapié en las tareas formales como juegos y demostración de teoremas. Sin embargo, como dicen E. Rich y K. Knight en su libro:

"Los investigadores de la IA se dedicaron inicialmente a estos problemas porque pensaban que en la solución de ellos se hallaba la esencia de la inteligencia. La visión o la comprensión del lenguaje (tareas que un niño de 5 años realiza correctamente) no eran consideradas problemas difíciles. Hoy día se dispone de programas de ajedrez tan potentes que están a la altura de los grandes maestros, pero no existe ningún programa que pueda llevar a cabo las características básicas de percepción que tiene un niño".

TEMA 1.- La manera natural de representar el mecanismo de un juego de tablero es mediante un árbol de juegos, que es un tipo especial de árbol semántico en el que los nodos representan configuraciones de tablero y las ramas indi-



can cómo una configuración puede transformarse en otra mediante un sólo movimiento.

Evidentemente en un juego como el ajedrez, una búsqueda exhaustiva en el árbol de juegos no puede utilizarse, dada la profundidad del árbol de ramificación, y se debe recurrir a estrategias de tipo heurístico. Para elegir el mejor movimiento, deben compararse las posiciones del tablero resultante para determinar cual es la más ventajosa. Esto se lleva a cabo con una función de evaluación estática.

La clave de un buen programa de juegos está en la elección de una adecuada función de evaluación estática, junto con un buen generador de movimientos. Gran parte del trabajo realizado en programas de juegos se ha efectuado desarrollando funciones de este tipo: Turing, Bernstein, Minsky...

En el procedimiento de búsqueda se requiere que se analicen tantos movimientos como sea posible. De la gran cantidad de algoritmos de búsqueda existentes, para una gran variedad de aplicaciones dentro de la IA, se han de escoger aquellos que sean adecuados para juegos bipersonales como el procedimiento de búsqueda minimax, debido a Claude Shannon (1950), donde un jugador espera obtener números positivos en la función de evaluación (jugador de maximización), y el otro jugador intentará forzar el juego hacia situaciones fuertemente negativas en la misma función (jugador de minimización).

La alternancia de maximización y minimización cuando las evaluaciones se envían a la raíz, se corresponden con las estrategias opuestas que siguen los dos jugadores y que da nombre a este método.

El rendimiento del algoritmo puede mejorarse incluyendo técnicas de ramificación y acotación.

En el momento de redactar estas líneas, Febrero de 1996, se está celebrando en Filadelfia, con motivo del 50 aniversario de la Association for Computing Machinery (ACM), un torneo de ajedrez a seis juegos entre Gary Kasparov y la máquina IBM's Deep Blue. Esta máquina es el resultado de un proyecto de IBM que comenzó en 1989, y cuyo objetivo era explorar la utilización del procesamiento paralelo para resolver problemas complejos.

El predecesor de Deep Blue fue Deep Thought, creado en la Universidad de Carnegie-Mellon en 1988, un sistema constituido por 2 procesadores y que era capaz de analizar 750.000 posiciones por segundo, situándose en el extremo inferior de la escala de los grandes maestros de ajedrez.

En octubre de 1989, una versión de Deep Thought, constituida por 6 procesadores y capaz de buscar más de 2 millones de posiciones por segundo, realizó una exhibición contra Gary Kasparov y perdió. En agosto de 1993, venció a Judit Polgar la más joven gran maestra de la historia. En junio de 1994, la computadora ganó el título de campeón internacional de ajedrez con computador.



El Deep Blue es un computador IBM Power Parallel SP2 con 32 nodos. Cada nodo emplea una tarjeta que contiene 8 procesadores VLSI específicos para ajedrez. El resultado es un computador con 256 procesadores, capaz de calcular de 50 a 100 miles de millones de movimientos en 3 minutos, que es el tiempo permitido a cada jugador de ajedrez en una partida típica. Además, el sistema incorpora una base de datos que incluye partidas realizadas con grandes maestros en los últimos 100 años. Si añadimos el hecho de que no se cansa o se distrae, ni se encuentra sometido a un esfuerzo psicológico, la computadora Deep Blue puede ser un serio (digno) contrincante de Kasparov.

TEMA 2.- Como indicamos al principio de este apartado, uno de los primeros dominios donde se aplicó la IA fue en la demostración automática de teoremas. Por ejemplo, el Logic Theorist (Newell, et al., 1963), que demostraba teoremas del Principia Mathematica de Whitehead y Russell (1950); el demostrador de teoremas (Gelernter, et., 1963), que trabajaba sobre teoremas de geometría ...

Aunque la demostración de teoremas sigue siendo un área de investigación activa en IA, sin embargo, su utilidad se ha extendido más allá del ámbito tradicional de las matemáticas. En efecto, la lógica (Boole elaboró su modelo algebraico de la lógica de proposiciones y Fregge formalizó la de predicados) que ha sido la herramienta que se ha utilizado para tratar de formalizar el razonamiento y permitir la demostración de teoremas, puede utilizarse como un medio para la representación del conocimiento. Para resolver los problemas complejos con los que se enfrenta la IA, es necesario disponer tanto de una gran cantidad de conocimiento como de una serie de mecanismos que permitan manipularlo con el fin de obtener soluciones a nuevos problemas. En los programas de IA se ha representado el conocimiento de muy distintas maneras. Una característica común a todas las representaciones es el de disponer de dos entidades:

- Hechos: verdades en un cierto mundo. Es aquello que queremos representar.
- Representaciones de los hechos en un determinado formalismo. Estas son las entidades que realmente seremos capaces de manipular.

Un buen sistema de representación del conocimiento, en un dominio particular, debe poseer las siguientes propiedades:

- 1.- Suficiencia en la representación
- 2.- Suficiencia deductiva
- 3.- Eficiencia deductiva
- 4.- Eficiencia en la adquisición.



La 1 como la capacidad de representar todos los tipos de conocimiento necesarios en el dominio. La 2 o la capacidad de manipular estructuras de la representación, con el fin de obtener nuevas estructuras que se correspondan con el conocimiento deducido. La 3 como la capacidad para incorporar información adicional en las estructuras del conocimiento, y finalmente la 4 para que la adquisición de una nueva información se realice con facilidad.

Como no se ha conseguido encontrar ningún sistema que optimice todos los aspectos mencionados, existen múltiples técnicas para la representación del conocimiento. Citaremos las más importantes:

- A.- Lógica de predicados de primer orden
- B.- Lógica de proposiciones
- C.- Representación del conocimiento mediante reglas
- D.- Razonamiento simbólico bajo incertidumbre: lógicas para razonamientos no monótonos
- E.- Razonamiento probabilístico: lógica bayesiana y lógica difusa.

Un predicado es la formalización de una propiedad. La construcción formada por un predicado, seguido de sus argumentos se denomina fórmula atómica. Una sentencia estará formada por el enlace con conectivas de fórmulas atómicas.

Se ha demostrado que no existe un procedimiento general que permita determinar la validez de cualquier sentencia del cálculo de predicados. Por ello, se dice que el cálculo de predicados es indecidible. Sin embargo, sí existe un procedimiento tal que si una sentencia es válida termina dictaminándolo y si no lo es, no para, por lo que también se dice que el cálculo de predicados es semi-decidible.

El atractivo de la lógica de proposiciones es su simplicidad. Es muy sencillo representar los hechos del mundo real como proposiciones lógicas, es decir, se dispone de un buen mecanismo para razonar con este conocimiento, aunque es computacionalmente más costoso que la lógica de predicados.

Un sistema axiomático se construye sobre la base de unos axiomas o construcciones que se admiten como verdaderas. Para definir un sistema axiomático hay que especificar: un alfabeto, las reglas de formación que permiten derivar sentencias, los axiomas y las reglas de transformación que permiten obtener o demostrar ciertas sentencias a partir de los axiomas.

El sistema axiomático más conocido es el denominado PM, nombre derivado del título de una obra clásica "Principia Mathematica" de Whitehead y Russell.

Una cláusula de Horn es una forma normalizada conjuntiva, donde cada uno de los operandos representa predicados que se aplican sobre argumentos, que tie-



nen como mucho un literal positivo. Cualquier expresión lógica puede convertirse en una cláusula de Horn. A diferencia de la lógica de predicados de primer orden, la lógica de sistemas de cláusulas de Horn es decidible.

Como lenguajes de programación lógica podemos citar en primer lugar LISP, creado por McCarthy, que junto con FORTRAN son los lenguajes más antiguos existentes, y es un lenguaje de procesamiento de listas. Los modelos paralelos de LISP incluyen MULTILISP y QLISP (1988).

PROLOG fue creado en 1975 por Colmerauer y Rousell para la demostración de teoremas, utiliza para las expresiones lógicas las cláusulas de Horn, y dispone en la actualidad de modelos paralelos como PARLOG (1986).

Supongamos ahora que una expresión es un teorema con respecto a cierto conjunto de axiomas. Después de añadir algunos axiomas nuevos, ¿Es la expresión todavía un teorema?. Con certeza debe serlo, ya que es posible hacer una demostración exclusivamente con los axiomas anteriores, haciendo caso omiso de los nuevos. Como los nuevos axiomas sólo contribuyen a aumentar la lista de los nuevos teoremas demostrables, y nunca propician la eliminación de cualquiera de éstos, se dice que la lógica tradicional es monótona.

Sin embargo, ciertos razonamientos humanos de sentido común son no monotónicos. Alcanzamos conclusiones a partir de ciertas premisas que no alcanzaríamos si otras sentencias se incluyeran entre nuestra premisas.

El razonamiento no monotónico, propuesto por primera vez a finales de la década de los años 70, se encuentra en rápido desarrollo. Se han propuesto diferentes alternativas para el soporte del razonamiento no monotónico: la lógica no monótona NML de McDermott y Doyle (1980), la lógica por defecto DL de Reiter (1980), razonamiento minimalista, la circunscripción de McCarthy,...

La representación del conocimiento mediante reglas y la lógica bayesiana se analizarán en los puntos correspondientes a los Sistema Expertos y al Reconocimiento de Patrones, respectivamente.

Para finalizar este apartado, dedicaremos unas líneas a una lógica que está teniendo actualmente una amplia acogida en multitud de aplicaciones. Nos estamos refiriendo a la lógica difusa de Zadeh.

Mientras que la teoría de conjuntos tradicional define el ser miembro de un conjunto como un predicado booleano:

$ACU, x \in U$, hay dos posibilidades $x \in A$ ó $x \notin A$.

La teoría de conjuntos difusa permite definir una función característica de pertenencia μ_A tal que $\mu_A(x)=1$ si $x \in A$ y $\mu_A(x)=0$ si $x \notin A$.

La pertenencia imprecisa del elemento x al conjunto A se expresa con un valor de $\mu_A(x)$ comprendido entre 0 y 1.



Se definen las operaciones de complementación, intersección, unión, de la siguiente forma:

$$\begin{aligned} \text{Complementación: } A = \bar{B} & \text{ si } \mu_A(x) = 1 - \mu_B(x), \forall x \in U \\ \text{Intersección} & : C = A \cap B \text{ si } \mu_C(x) = \min(\mu_A(x), \mu_B(x)), \forall x \in U \\ \text{Unión} & : C = A \cup B \text{ si } \mu_C(x) = \max(\mu_A(x), \mu_B(x)), \forall x \in U \end{aligned}$$

La forma de las sentencias en lógica borrosa o difusa es la misma que en lógica de predicados, sin embargo, para cada sentencia se tiene que poder calcular un significado, es decir, un subconjunto difuso que corresponda a la interpretación "verdadera" de la sentencia.

TEMA 3.- Los Sistemas Expertos (SE) constituyen la mayor parte de la IA aplicada y resuelven problemas que normalmente son solucionados por "expertos humanos". Estos problemas resueltos por los SE son los típicos realizados por especialistas, con un saber acumulado por muchos años de experiencia, y en el que gran parte de éste reside en su subconsciente. Por ello, una de las labores más complejas a las que se enfrenta un ingeniero del conocimiento es extraer el conocimiento de ese experto.

La forma más utilizada de representación del conocimiento en los SE es la representación mediante reglas de antecedente-consecuente en la forma "if then else", es decir, a los patrones del si (condiciones) se les denomina antecedentes, mientras que a los patrones de la palabra clave entonces se les denomina consecuentes.

Pongamos como ejemplo el SE MYCIN, un sistema experto para la determinación de enfermedades infecciosas bacterianas. Una regla puede ser de la forma:

IF (1) la infección es debida a bacterias primarias, y
 (2) la localización del cultivo es una de las muestras estériles, y
 (3) la entrada probable es el tracto gastrointestinal,
 THEN Es bastante probable (0.7) que la identidad de los organismos sea bacteriana.

donde se han utilizado además factores de certeza.

Los componentes principales de un SE son:

- la base de conocimientos, que contiene el conocimiento de los hechos y la experiencia de los expertos expresada en reglas.
- el mecanismo de inferencia, que simula la estrategia del experto extrayendo conclusiones de la base de conocimientos.



Existen dos direcciones de funcionamiento en el mecanismo de inferencia:

- razonamiento hacia delante a partir de los estados iniciales
- razonamiento hacia atrás a partir de los estados objetivo.

En los SE se suelen utilizar uno de los dos razonamientos o una combinación de ambos. Por ejemplo, MYCIN utiliza razonamiento hacia atrás para detectar organismos que están presentes y razonamiento hacia delante para razonar sobre el regimen de tratamiento.

Las aplicaciones de los SE en medicina requieren que éstos nos informen de cual ha sido el camino que ha seguido para producir un diagnóstico. Por ello, se exige a los SE actuales que dispongan del conocimiento necesario para que nos informen de su propia gestión (metaconocimiento).

Otra característica que se exige a los actuales SE es el aprendizaje, como fue definido por Simon en 1983

"...cambios en el sistema que le permitan llevar a cabo la misma tarea o tareas a partir de las mismas condiciones de un modo más eficiente y eficaz cada vez"

Entre los diferentes métodos de aprendizaje destaca el realizado por medio de ajuste de parámetros. En este método el programa utiliza una función de evaluación estática $c_1t_1+c_2t_2+\dots+c_nt_n$, donde t_i es una característica que contribuye a la evaluación. El término c_i es el peso que se asigna a cada característica. A medida que el aprendizaje progresa, el valor de los pesos va cambiando. Aquellos coeficientes de los términos cuya capacidad de predicción es muy pobre deben disminuir, mientras que los coeficientes de los términos que predicen el final deben aumentar.

En un principio los SE se construían generalmente en LISP. Desde que los SE se realizan como un conjunto de representaciones declarativas combinadas con un intérprete para esas declaraciones, fue posible separar el intérprete y el conocimiento específico del dominio. De este modo, se pudo crear un sistema capaz de construir nuevos SE sin más que añadir nuevos conocimientos correspondientes al nuevo problema. Los intérpretes resultantes se denominan armazones o shells. Podemos citar como armazones más conocidos ART Enterprise, Nexpert,...

Un requerimiento que imponen algunas aplicaciones es que los SE puedan actuar en tiempo real, es decir, que la respuesta del mismo se disponga en el tiempo requerido, independientemente del número de reglas que tengan que dispararse en cada momento. Para ello, la estrategia del SE debe cambiarse para que



la respuesta se obtenga desde el primer momento, mejorándose en instantes sucesivos.

Cuando ésto no es posible realizarlo en un sistema monolítico, se recurre a los sistemas con razonamiento distribuido, que son sistemas compuestos por un conjunto de módulos, llamados agentes, y un conjunto de caminos de comunicación entre ellos, necesitándose para su implementación una arquitectura de tipo paralelo formada por varios procesadores.

Existen multitud de situaciones en las que resulta apropiada la utilización de los SE: Prospecciones petrolíferas, medicina, economía, elección de sistemas informáticos (Digital Equipment), detección de fallos en circuitos integrados,... pero es sobre todo en medicina donde existe una mayor variedad: diagnóstico de patologías, en unidades de cuidados intensivos, selección de fármacos,...

TEMA 4.- El objetivo del reconocimiento de patrones es reconocer las diferentes versiones de un objeto. Una parte esencial en este reconocimiento lo constituye la clasificación, entendida como la asignación de objetos a clases predefinidas, en base a sus características. Podemos dividir la clasificación en dos etapas: extracción de patrones² y la propia clasificación. La fase de extracción de patrones trata de filtrar la información sobre el objeto a clasificar, de manera que se seleccionen las características relevantes al problema a partir de las propiedades observadas del objeto. En la segunda fase, se realiza la clasificación en base a esas características elegidas. Ante la imposibilidad de definir un modelo determinista de las clases éstas deben de ser caracterizadas por medio de las densidades de probabilidad multivariantes de las características observadas. Surgen aquí los distintos tipos de clasificadores estadísticos, según estas distribuciones de probabilidad se estimen de la muestra o se hagan suposiciones sobre la forma de las mismas.

Las funciones discriminantes asociadas a cada clase y que utilizará el clasificador para asignar un objeto a una de ellas, se determinan a partir de las probabilidades condicionales de pertenencia a cada clase, conocido el valor de las observaciones. A través de la ley de Bayes, estas probabilidades condicionales pueden calcularse a partir de la probabilidad de la observación, supuesto que el objeto pertenece a una clase determinada y que normalmente se aproximan a gaussianas, es decir, cada clase vendrá caracterizada por su vector de medias y su matriz de covarianza.

Las redes neuronales artificiales han mostrado ser una alternativa válida a los clasificadores de tipo estadístico, donde no se requiere hacer ningún tipo de

² Denominamos patrón al conjunto de características de tipo cuantitativo que pueden extraerse de un objeto.



suposición estadística. Estudios recientes han demostrado el mejor comportamiento en la clasificación de las redes de tipo perceptrón multicapa que los clasificadores estadísticos, en un problema de determinación de la maduración cerebral en niños a partir de señales electroencefalográficas.

De la misma manera que ha sido aplicada la IA al reconocimiento de patrones, puede también aplicarse a otro campo solapado con éste como es el reconocimiento y tratamiento de imágenes en procedimientos de análisis de imágenes de satélites o imágenes médicas, donde se requieren procesos de clasificación de objetos en tres dimensiones.

TEMA 5.- Si atendemos a la definición de IA de Pamela McCorduck, dada al principio de este trabajo, o a la afirmación de John McCarthy sobre la inteligencia de los termostatos, no cabe ninguna duda de la enorme influencia que ha tenido la Automática en la IA, por su contribución a la creación de autómatas en la segunda mitad del siglo XX, y sobre todo del autómata por antonomasia: el robot.

El principio fundamental de la Automática o Control es la realimentación para reducir la influencia de las perturbaciones que actúan sobre el sistema a controlar. El proceso de control se realiza a través de la percepción, el "razonamiento" y la actuación, a intervalos regulares de tiempo, cuyo periodo de muestreo depende de los requerimientos del sistema.

La poderosa herramienta matemática que ha sido creada durante los últimos 50 años, referida a los sistemas dinámicos lineales y no lineales, continuos y discretos multivariados, ha permitido el control de complejos sistemas: desde una planta química, los vuelos espaciales, hasta el robot ³.

Esta herramienta ha desembocado en lo que actualmente denominamos Teoría de Sistemas o Sistémica, que está teniendo una influencia creciente en campos tales como la economía, la biología, la psicología,...

Los Sistemas Expertos en tiempo real, la Lógica difusa y las redes neuronales están siendo aplicados en la actualidad con indudable éxito en el control de procesos.

En la Robótica⁴, donde estas máquinas desempeñan actividades monótonas, repetitivas y en muchos casos peligrosas, se intenta que su comportamiento sea cada vez más inteligente, además de que dispongan de una mayor capacidad

³ no mencionaremos aquí los enormes avances realizados en la industria armamentística promovidos por la Automática

⁴ La palabra Robot se introdujo en el inglés a partir de la obra de Karel Capek R. U. R. (1923). Proviene de la palabra eslava robota, que significa trabajos forzados.



visual y sensora. Las redes neuronales estáticas se están aplicando en la realización de controladores, y recientemente, las redes neuronales dinámicas se están utilizando en la modelización de la dinámica de los robots.

TEMA 6.- La tarea de investigación de la IA en el procesamiento del lenguaje natural estriba en disponer en la máquina de capacidades ligadas a la comprensión inteligente del lenguaje, y facilitar la comunicación entre el hombre y la máquina.

Un sistema de procesamiento de lenguaje natural es aquél en que:

- parte de la información a procesar está codificada en lenguaje natural y
- se aplican algoritmos para el análisis sintáctico, semántico y pragmático de la información y para la generación en lenguaje natural.

Uno de los más célebres programas de IA fue el sistema ELIZA, realizado por Joseph Weizenbaum en 1966, y simulaba el comportamiento de un psicoterapeuta, aunque ELIZA era en realidad un programa para el estudio de la comunicación en lenguaje natural entre hombre y máquina, tal como lo definió su autor. Según P. McCorduck

"el programa era impresionante, aunque su mecanismo era relativamente sencillo, creaba la ilusión de que realizaba un análisis semántico a fondo. ELIZA generaba respuestas en lenguaje natural a las frases que recibía en lenguaje natural. Cuando algo le desconcertaba, ELIZA podía recurrir a muletillas como "ya veo", o "eso es muy interesante", o "continúe", igual que hacen los humanos en situaciones similares".

T. Winograd, profesor de Stanford, adoptó como hipótesis de trabajo que el conocimiento sintáctico y semántico no podían separarse de los hechos, que el significado está incorporado en estructuras procedimentales. Su programa denominado SHRDLU, realizaba el razonamiento, el análisis semántico y el análisis sintáctico simultáneamente en lugar de secuencialmente. Situaciones de ambigüedad semántica, que en anteriores programas llegaban a callejones sin salida, con este programa no se producían.

Posteriormente, se hicieron mejoras de SHRDLU, dando lugar a una nueva versión denominada HACKER, la cual aprendía de sus propios errores.

Noam Chomsky mantenía que cada uso del lenguaje se refería a una estructura subyacente de conceptos y datos, y que el significado era un conjunto de operaciones realizadas sobre la base de esa estructura.



Roger Schanck, profesor de Stanford, enfocó el problema de forma similar a Chomsky, creando la idea de guión muy similar a la de marco de Minsky. Tanto el guión como el marco sirven de esqueleto a partir del que construimos nuevas asociaciones.

Después de evidentes avances realizados en este campo, aún quedan cuestiones básicas de las que sólo se tienen respuestas muy vagas:

"¿Hasta qué punto es esencial la comprensión para el reconocimiento del habla?.

¿Cómo aprendemos tanto nosotros como las computadoras en situaciones realmente nuevas, si la mayor parte de nuestro aprendizaje se da por comparación del presente con el pasado y por el recurso –y la depuración– de antiguas rutinas para hacer que se adecúen a las situaciones nuevas?

¿Y cómo adquirimos esas viejas rutinas la primera vez?

Si lo que impide que tomemos un camino equivocado (en el caso de ambigüedades, por ejemplo) es algún mecanismo de aviso que se dispara debido a la experiencia o a conocimiento adquirido plenamente, ¿Por qué los mecanismos de aviso (conocidos en IA como demonios) funcionan en algunos agentes inteligentes mejor que en otros?

¿Podemos aprender a ser más inteligentes?

¿Podemos organizar el conocimiento de forma más eficiente en nuestras cabezas?"

En fin, todas estas preguntas, planteadas por P. McCorduck nos dan una idea de la envergadura del problema que la IA pretende resolver, y que le hicieron decir a Raj Reddy, profesor de Stanford, lo siguiente:

"Si eres muy calculador, muy cuidadoso, una persona sistemática y racional que sopesa el más mínimo proyecto que cae en tus manos para ver si funcionará, y seleccionar sólo cosas que son seguras, entonces obviamente no puedes dedicarte a la IA".

LA HISTORIA DE LA IA

La IA a lo largo de su corta historia ha tenido dos escuelas o tendencias: la IA conexionista y la IA simbólica. Ambas tendencias han luchado por su reconocimiento, la IA conexionista pretendía encontrar un modelo del cerebro, por lo que se consideraba como una neurociencia idealizada; la otra, la IA simbólica consideraba las computadoras como sistemas para manipular símbolos men-



tales. Cada una de ellas ha podido presentar importantes resultados y deberían considerarse no como dos modelos alternativos y rivales sino como planteamientos complementarios de una misma cosa: tratar de emular la mezcla de deseo e inteligencia que constituye gran parte de lo que piensan y hacen los seres humanos, sin embargo, ambas escuelas se han desarrollado dándose la espalda. Un cuento presentado por Seymour Papert en su artículo "¿Una sola IA o muchas?" explica muy claramente la controversia entre ambas.

El cuento dice así:

"Había una vez dos ciencias hermanas de la nueva ciencia de la Cibernética. Una hermana era natural y tenía rasgos heredados del estudio del cerebro, de la forma en que la naturaleza hace las cosas, la otra era artificial, relacionada desde el comienzo con el uso de las computadoras.

Cada una de las ciencias hermanas procuraba construir modelos de la inteligencia, pero a partir de materiales muy diferentes. La hermana natural construía modelos (llamados redes neuronales) con neuronas matemáticamente purificadas. La hermana artificial construía sus modelos con programas de computación.

En el primer florecer de su juventud ambas eran igualmente exitosas e igualmente codiciadas por pretendientes de otros campos del conocimiento. Juntas se llevaban bien. Pero sus relaciones cambiaron a principios de la década de los 60, cuando apareció un nuevo monarca, uno que poseía los cofres más grandes que se hubieran visto en el reino de las ciencias: El señor DARPA (Agencia de Proyectos de investigación Avanzadas del Departamento de Defensa de los EEUU). La hermana artificial se tornó celosa y tomó la decisión de guardar para sí el acceso a los fondos de investigación del señor DARPA. La hermana natural debía ser eliminada.

El trabajo sangriento fue intentado por dos seguidores de la hermana artificial Marvin Minsky y Seymour Papert puestos en el papel de cazador enviados a matar a Blancanieves y a traer su corazón como prueba. El arma no era una daga sino una pluma, mucho más poderosa, de la que salió un libro: Perceptrons, destinado a probar que las redes neuronales nunca podrían cumplir su promesa de construir modelos de la mente: sólo los programas de computación podrían hacerlo. Y ciertamente, durante la siguiente década todas las recompensas del reino fueron para su progenie, de la cual la familia de los sistemas expertos hizo la mejor fama y fortuna.

Pero Blancanieves no estaba muerta. Lo que Minsky y Papert habían mostrado al mundo como prueba de su muerte no era su corazón; era el corazón de



un cerdo. Para ser más literales: en el libro podía leerse la prueba de que la estrategia de las redes neuronales para construir modelos de la mente estaba muerta. El libro señalaba limitaciones muy serias de una cierta clase de redes (hoy en día conocidas como perceptrones de una sola capa); pero se encontraba equivocado en lo concerniente a la suposición de que esta clase de red era el corazón del conexionismo.

En aquel tiempo y lugar tan alejados, todavía estaban por hacerse los descubrimientos técnicos que abrirían la visión de redes neuronales mucho más poderosas de lo que entonces podía imaginarse.

En los viejos días de Minsky y Papert, los modelos de redes neuronales se hallaban limitados y sin esperanza por el encanijamiento de las computadoras disponibles en la época y por la falta de ideas sobre cómo hacer que una red que no fuera de las más simples aprendiera algo. Ahora las cosas han cambiado. Computadoras poderosas, masivamente paralelas, pueden implementar redes muy grandes y nuevos algoritmos pueden hacer que aprendan. No se necesita un Príncipe Encantador para esta historia.

Las escrituras conexionistas presentan la historia con un final feliz. la hermana natural había sido cuidada en los laboratorios de unos pocos investigadores ardientes que sostuvieron su fe, incluso cuando el mundo en general estaba convencido de que la empresa era fútil."

En la historia de la IA se pueden distinguir varias fases. Una primera de gestación entre 1943 y 1956, con un primer trabajo realizado en 1943 por McCulloch y Pitts, proponiendo un modelo de neuronas artificiales para el cerebro.

Donald Hebb propuso en 1949 que la conectividad del cerebro cambiaba continuamente, a medida que un organismo aprendía tareas funcionales diferentes.

A comienzos de los años 50, Shannon (1950) y Turing (1953) crearon los primeros programas de ajedrez.

Hacia 1955 Newell y Simon, trabajando en la Rand Corporation, llegaron a la conclusión de que unas cadenas de bits manipulados por una computadora podían representar cualquier cosa: números o rasgos del mundo real.

La siguiente fase en la evolución de la IA abarca los años que median entre 1956 y 1966. Fue ésta una etapa de grandes y prometedoras expectativas. Se abre con la conferencia del Dartmouth College.

En 1958 Rosenblatt concibió el perceptrón, que puede ser visto como una unidad lógica binaria capaz de implementar las tres operaciones del álgebra de Boole: Y, O y complementación, pero incapaz de implementar algunas funciones lógicas más complejas, por ejemplo XOR. Es un elemento constituido por n entradas sumadas de forma ponderada. El resultado se pasa a través de una no linealidad todo-nada. En los años 60 se propusieron diferentes algoritmos de aprendizaje para este perceptrón.



En ese mismo año McCarthy definió el lenguaje de alto nivel LISP, que ha sido un lenguaje predominante en IA. También en ese mismo año McCarthy concibió el "Advice Taker" que, como el GPS, utilizaba el conocimiento para buscar soluciones a problemas.

En 1959 Herbert Gelernter construyó el Demostrador de Teoremas Geométricos. Por estos mismos años Newell y Simon construyeron el "General Problem Solver" (GPS).

Poco después de las primeras publicaciones de Rosenblatt, apareció una variante del perceptrón, inventada por B. Widrow y M.E. Hoff, llamada ADALINE (Adaptive linear neuron), que se diferenciaba de aquella únicamente en el proceso de entrenamiento.

En 1965 Weizenbaum presentó el programa ELIZA.

A partir del año 1966, comienzan los problemas para la IA. Inicialmente, con el intento fallido de creación de traducción automática entre el ruso y el inglés que obligó a cancelar las subvenciones del gobierno para todos los proyectos de traducción automática.

El segundo tipo de problema apareció con la intratabilidad⁵ de los problemas que la IA pretendía resolver. Los procedimientos de búsqueda necesarios para resolverlos tenían que enfrentarse con la explosión combinatoria a que daban lugar.

El tercer problema surgió en 1969 con la aparición del libro "Perceptrons" de Minsky y Papert que demostraba que los perceptrones de Rosenblatt no podían distinguir entre patrones tan simples como la T y la C independientemente de las traslaciones y rotaciones que tuvieran lugar, lo que quería indicar que los perceptrones o adalines elementales no eran computacionalmente universales. Este hecho produjo realmente un parón en esta escuela de la IA, hasta que no se encontraron procedimientos de entrenamiento y estructuras de redes adecuadas, y esto no ocurrió hasta la década de los años 80.

La etapa que va desde 1966 hasta 1979 constituyó una cura de humildad para la IA. Fue una etapa de búsqueda de mecanismos adecuados para la representación del conocimiento. Además se desarrolló una nueva metodología: Los Sistemas Expertos.

El primero fue el SE DENDRAL, desarrollado por Buchanan, Feigenbaum y Lederberg en 1969 en Stanford, con el objetivo de deducir la estructura mole-

⁵ SE dice que un algoritmo tiene complejidad en el tiempo $O(f(n))$ si el número de pasos que se necesitan para procesar un dato de tamaño n es $cf(n)$, donde $f(n)$ es una función de n y c una constante. Se define un problema como tratable si y sólo si existe un algoritmo para resolverlo que tiene una complejidad en el tiempo $O(p(n))$, donde $p(n)$ es una función polinomial del tamaño de la entrada n .

Un problema en el que todos los algoritmos tienen una complejidad en el tiempo $O(kn)$, es decir, un tiempo de ejecución exponencial con n , es intratable.



cular de un complejo orgánico, a partir de la información suministrada por espectrógrafos de masas.

El siguiente esfuerzo de realización de un SE se realizó en medicina para el diagnóstico médico de infecciones en la sangre con MYCIN, el cual incorporaba razonamiento con incertidumbres.

En este periodo se realizó una importante aportación en el área de la comprensión del lenguaje natural con el SHRDLU de Winograd en el MIT, y Schank en Yale.

El crecimiento de las aplicaciones de la IA a problemas del mundo real exigió disponer de esquemas adecuados para la representación del conocimiento. Algunos estaban basados en la lógica como PROLOG (1975), otros como los marcos (frames) de Minsky (1975) supusieron disponer de un formalismo para la representación del conocimiento, que permitiera describir los objetos a través de un almacén estructurado de datos, que se iba rellenando a medida que se avanzaba en el procesamiento.

El último periodo de esta breve historia de la IA abarca desde 1980 hasta la actualidad y se caracteriza ,por un lado, por las aplicaciones industriales de los SE y ,por otro, por el retorno de las redes neuronales.

En efecto, se han desarrollado una gran variedad de SE en áreas muy diversas. A modo de resumen, indicaremos las siguientes:

- ámbito militar

.AALPS: optimización de carga en aviones.

- medicina

.ABEL: disfunciones metabólicas

.ATTENDING: anestesia preoperatoria

.CAA:Arritmias cardíacas

.BIRTH DEFECTS DIAGNOSIS: enfermedades congénitas

.PATRICIA: unidad de cuidados intensivos

.ONCOCIN: tratamiento terapéutico en enfermedades de cáncer

.SEDPOSC: diagnóstico de patologías observables en señales cerebrales.

- arte

.AARON: creación de gráficos abstractos.

- química

.CRYALIS: estructura tridimensional de proteínas

.DENDRAL: estructura molecular

- economía

.APEX: planificación financiera

.COMEX: bolsa



- .ESCORT: control de producción.
- enseñanza
 - .GUIDON/NEOMYCIN: cualidades diagnósticas
 - .KLAUS: enseñanza oral.
- derecho
 - .LEGOL: jurisprudencia.
- electrónica
 - .DART: apoyo en el diseño de circuitos integrados
 - .EL: análisis de circuitos eléctricos
 - .PHOTOLITHOGRAPHY ADVISOR: diagnóstico de fallos en la producción de circuitos integrados

No podíamos terminar aquí la evolución de la rama simbólica de la IA sin mencionar los esfuerzos realizados con el conocimiento basado en el sentido común, en el que un ejemplo muy importante es el razonamiento basado en casos (RBC).

Hasta ahora los programas de IA han resuelto los problemas razonando a partir de primeros principios, sin embargo, los sistemas RBC resuelven los nuevos problemas por analogía con los problemas antiguos ya resueltos.

CYC (Lenat y Guha, 1990) es un proyecto de una gran base de conocimiento, cuyo objetivo es la captura del conocimiento humano de sentido común. Utiliza tanto los marcos como la lógica para codificar los tipos específicos de conocimiento.

El objetivo de esta gran base de conocimiento sería el de combinarse con bases de conocimiento especializadas para producir sistemas menos frágiles que los actuales⁶, en el sentido de que presionados contra los bordes no se quiebran, es decir, que no den respuestas absurdas a preguntas que estén ligeramente fuera de los dominios para los que fueron programados.

Ya indicamos que la aparición del libro "Perceptrons" produjo un auténtico parón en la línea conexionista de la IA. Sin embargo, psicólogos como David Rumelhart y Geoff Hinton continuaron el estudio de los modelos de redes neuronales. Físicos como Hopfield utilizaron técnicas de mecánica estadística para analizar las propiedades de almacenamiento y optimización de las redes. La línea conexionista resurgió cuando se generalizaron los perceptrones a redes

⁶ contruidos a partir del paradigma estándar de la IA: elegir un problema significativo en que todo el mundo esté de acuerdo en que es necesario para resolverlo utilizar la inteligencia; identificar los elementos de información que se necesitan para lograr una solución del problema; determinar cómo podría representarse esa información en una computadora; encontrar un algoritmo que pueda manipular esta información; escribir un código computacional que implemente ese algoritmo y ponerlo a prueba contra instancias muestreadas del problema.



neuronales de múltiples capas y se encontraron los algoritmos de entrenamiento de propagación hacia atrás. Estas técnicas se popularizaron en la colección "Parallel Distributed Processing" de Rumelhart y McClelland en 1986. Desde entonces, estos algoritmos, junto con otros más orientados hacia aprendizaje no supervisado, han sido aplicados a problemas en una gran variedad de áreas: economía, control, robótica, reconocimiento de patrones, ... con un éxito sorprendente.

Los sistemas simbólicos no está claro que puedan resultar adecuados si el contexto cambia, sin embargo las redes neuronales no precisan de una descripción rigurosa sino una actualización de los pesos a la nueva situación y, por tanto, tienen un comportamiento relevante ante cambios de contexto.

He pretendido en este trabajo presentar la disciplina de la IA: su definición, los temas que trata y su historia. A lo largo del trabajo, he tratado de dar algunas de las opiniones de los más prestigiosos investigadores que han contribuido con sus trabajos a engrandecerla. Por otra parte, he tratado de mantenerme, como no podía ser de otra manera, al margen y presentar las diferentes corrientes de la forma más objetiva posible, sin pretender desnivelar la balanza en ningún sentido, introduciendo un sesgo interesado.

Podemos concluir el trabajo con la siguiente afirmación de Cowan y Sharp:

"Predecimos que la estrategia de arriba hacia abajo de la IA convencional y la estrategia de abajo hacia arriba del conexionismo pueden unirse eventualmente para producir un progreso real en lo que McCulloch una vez llamó EPISTEMOLOGÍA EXPERIMENTAL: el estudio de la forma en que el conocimiento se encarna en el cerebro y podría encarnarse en las máquinas".

**BIBLIOGRAFÍA.**

- 1) Dieter Nebendahl "Sistemas Expertos". Marcombo Boixareu Editores. 1988.
- 2) Roger Penrose "La nueva mente del Emperador". Mondadori. 1991.
- 3) Gregorio Fernández, Fernando Saez Vacas "Fundamentos de informática" Alianza Editorial. 1987.
- 4) S. J. Russell, P. Norving "Artificial Intelligence". Prentice Hall. 1995
- 5) Stephen R. Graubard (com.) "El nuevo debate sobre la Inteligencia Artificial". Gedisa Editorial. 1988.
- 6) Pamela McCorduck "Máquinas que piensan". Tecnos. 1991.
- 7) John Searle "Mentes, Cerebros y Ciencia". Segunda Edición. Ediciones Cátedra.1990.
- 8) Bruce Mazlish "La cuarta discontinuidad. La coevolución de hombres y máquinas". Alianza Editorial. 1994.
- 9) Richard L. Gregory (ed.) "Diccionario Oxford de la mente".Alianza Editorial. 1995.
- 10) Geoffrey E. Hinton, David C. Plant, Tim Shallice "Simulating Brain Damage". Scientific American. October. 1993
- 11) Geoffrey E. Hintor "Redes Neuronales que aprenden de la experiencia" Investigación y Ciencia. Noviembre. 1992.
- 12) John R. Searle "¿Es la mente un programa informático? Investigación y Ciencia. Marzo. 1990.
- 13) Paul M. Churchland, Patricia Smith Churchland "¿Podría pensar una máquina? Investigación y Ciencia. Marzo. 1990.
- 14) Don R: Hush, Bill G. Horne "Progress in Supervised Neural Networks". IEEE Signal Processing Magazine. January. 1993.



- 15) J. E. Hopcroft, J. D. Ullman "Formal languages and their relation to automata". Addison Wesley. 1969.
- 16) William R. Ashby "Proyecto para un cerebro". Editorial Tecnos. 1965.
- 17) Elaine Rich, Kevin Knight "Inteligencia Artificial" Segunda Edición. Mc Graw Hill. 1994.
- 18) Patrick Henry Winston "Inteligencia Artificial". Tercera Edición. Addison-Wesley Iberoamericana. 1994.
- 19) T.O' Shea, M. Eisenstadt "Artificial intelligence: Tools, Techniques and Applications". Harper and Row Publishers. 1984.
- 20) <http://chess.ibm.park.org/>
<http://chess.ibm.park...ep/blue/dbphotos.html>
- 21) L. Moreno, J. D. Piñeiro, J. L. Sánchez, S. Mañas, J.J. Merino, L. Acosta, A. Hamilton. "Using Neural Networks to Improve Classification: Application to Brain Maturation". Neural Networks, Vol. 8, Núm. 5. 1995. Elsevier Science.
- 22) L. Moreno, L: Acosta, A. Hamilton. "Neural Network Based Minimum-Time Optimal Control of Nonlinear System". World Congress on neural Networks Vol. III. Portland. 1993.
- 23) Javier Aracil "Introducción a la Dinámica de Sistemas". Alianza Universidad. Tercera edición. 1992.
- 24) L. Acosta, G. N. Marichal, A. Hamilton, J. A. Méndez, L. Moreno. "Design of a Dynamical Neural Network Structure for System Identification". International Computer Science Convention ICSC'96, Intelligence Industrial Automation. University of Reading. U.K. March. 1996.
- 25) L. Moreno, J.D. Piñeiro, R. M. Aguilar, J.L. Sánchez, S. Mañas, J. J. Merino, J. I. Estevez "Knowledge Based System for Diagnosis Observable in Brain Signals" IEEE International Conference on Systems, Man and Cybernetics. Canada. October. 1995.

Los autores con los artículos que aparecen en el texto, y que no figuran en esta bibliografía, corresponden a artículos recogidos en la referencia 5.