**AI: HISTORIA TEMPRANA Y APLICACIONES**

*Todos los hombres por naturaleza desean saber...*

-ARISTOTELES*,* Frase de apertura de la *Metafísica*

*Escuchar el resto, y se maravilla aún más a la artesanía y recursos que he ideado.*

*Mayor era la siguiente: en los tiempos antiguos, si un hombre cayó enfermo que no tenía ninguna defensa contra el la enfermedad, la curación ni alimentos ni bebidas, ni ungüento;* *sino por la falta de medicamentos hombres consumido, hasta que les mostré la mezcla de simples suaves con que expulsan toda clase de enfermedades.* *.* *.* *.*

*Fue lo que hizo visible a los ojos de los hombres de los signos de fuego del cielo que han sido antes tenue. Esto en cuanto a éstos.* *Por debajo de la tierra, del hombre oculto bendición, cobre, hierro, plata, y oro-reclamará a nadie que éstos han descubierto antes que yo?* *Nadie, estoy muy seguro, que quiere hablar de verdad y para el propósito.* *Una palabra breve va a contar toda la historia: todos artes que los mortales han venido de Prometeo.*

-AESCHYLUS, *Prometeo encadenado*

**1.1 De Edén, a ENIAC: Las actitudes hacia la inteligencia, el conocimiento, y el artificio humano.**

Prometeo habla de los frutos de su transgresión contra los dioses del Olimpo: su propósito no era simplemente robar el fuego para la raza humana, sino también para iluminar a la humanidad a través del don de la inteligencia o *nous:* la *mente racional.* Esta inteligencia forma base de toda la tecnología humana y, en última instancia, toda la civilización humana. El trabajo de Esquilo, el dramaturgo griego clásico, ilustra una conciencia profunda y antigua del extraordinario poder del conocimiento. La inteligencia artificial, en su interés muy directo en el don de Prometeo, se ha aplicado a todas las áreas de su legado, la medicina, la psicología, la biología, la astronomía, la geología y muchas áreas de la investigación científica que pudiera Esquilo no han imaginado.

Aunque la acción de Prometeo de liberar a la humanidad de la ignorancia de la enfermedad, que también le valió la ira de Zeus. Indignado por este robo de conocimiento que previamente pertenecía sólo a los dioses del Olimpo, Zeus ordenó que Prometeo sea encadenado a una roca estéril a sufrir los estragos de los elementos para la eternidad. La idea de que los esfuerzos humanos para adquirir conocimientos constituyen una transgresión contra las leyes de Zeus o la naturaleza está profundamente arraigada al pensamiento occidental. Es la base de la historia del Edén y aparece en la obra deDante y Milton. Tanto Shakespeare como las Antiguas tragedias griegas representan la ambición intelectual como causa de desastres. La creencia de que el deseo de conocimiento debe en última instancia conducir al desastre ha persistido a lo largo de la historia, que dura la época de la Ilustración e incluso por los avances científicos y filosóficos de los siglos XIX y XX. Así que no debe sorprender el hecho de que la Inteligencia Artificial inspirar tanta controversia en los círculos académicos y populares.

De hecho, en lugar de disipar este temor antiguo de las consecuencias de la ambición intelectual, la tecnología moderna ha hecho estas consecuencias sólo se parecen más probable, incluso inminente. Las leyendas de Prometeo, Eva Fausto se volvieron a contar en el lenguaje de la sociedad tecnológica. En la introducción de Frankenstein, con interesantes subtitulado *El Moderno Prometeo*, Mary Shelley escribe:

Muchos y largas eran las conversaciones entre Lord Byron y Shelley, que era un agente de escucha devota y silenciosa. Durante una de las, varias enseñanzas discutidas en forma filosófica, y entre otras, la naturaleza es el principio de la vida, y si había alguna posibilidad de que incluso sea cubierta y comunicada. Hablaron de los experimentos del Dr. Darwin (No me refiero a lo que el médico realmente hizo o dijo que sí, pero como mi objetivo, mientras hablaba de que él habría hecho), que conservan un pedazo de vidrio hasta que, por algunos medios extraordinarios, se empezó a mover voluntariamente. Al final, no es de esta manera que la vida debe ser creada. Tal vez un solo cuerpo se revivieron; galvanismo señaló que: tal vez las partes componentes de una criatura pueden ser fabricadas, montadas y equipadas con calor vital (Butler, 1998).

Mary Shelley nos muestra en qué grado los avances científicos tales como el trabajo de Darwin y el descubrimiento de la electricidad habían convencido de que incluso los no científicos funcionamiento de la naturaleza no eran secretos divinos, pero podría dividirse y entendidos de manera sistemática. El monstruo de Frankenstein no es el producto de encantamientos chamánicas o transacciones inaudibles como un submundo: y montado de componentes separadamente “fabricados" e infundidos con la fuerza vital de la electricidad. Aunque la ciencia del siglo XIX era insuficiente para alcanzar el objetivo de la comprensión y la creación de un agente totalmente inteligente, afirmó la idea de que los misterios de la vida y la inteligencia pueden ser llevados a la luz del análisis científico.

* + 1. **Una breve historia de los Fundamentos para la IA**

Cuando Mary Shelley finalmente, y tal vez irrevocablemente, se unió a la ciencia moderna con el mito de Prometeo, los fundamentos filosóficos de trabajo moderno en inteligencia artificial habían sido desarrollados durante varios miles de años. A pesar de los problemas morales y culturales planteados por la inteligencia artificial son a la vez interesantes e importantes, nuestra introducción se preocupa más con la herencia del patrimonio intelectual de IA. El punto de partida lógico para una historia de este tipo es el genio de Aristóteles, o como Dante en la Divina Comedia se refiere a él, "el maestro de los que saben". Aristóteles se reunió con los descubrimientos, las maravillas, y los temores de la antigua tradición griega con el análisis cuidadoso y pensamiento disciplinado que se han convertido en el estándar para la mayoría de la ciencia moderna.

Para Aristóteles, el aspecto más fascinante de la naturaleza era el cambio. En su Física, él define su "filosofía de la naturaleza" como el "estudio de las cosas que cambian". El distinguió entre la *materia* y la *forma* de las cosas: una escultura está formada de material de bronce y tiene la forma de un ser humano. El cambio ocurre cuando el bronce se moldea a una nueva forma. La distinción materia / forma proporciona una base filosófica para las nociones modernas como la computación simbólica y la abstracción de datos. En el cómputo (incluso con números) estamos manipulando los patrones que son las formas de material electromagnético, con los cambios de forma de este material que representa aspectos del proceso de solución. Abstraer la forma del medio de su representación no sólo permite que estas formas sean manipuladas computacionalmente, sino que también proporciona la promesa de una teoría de estructuras de datos, el corazón de la informática moderna. También es compatible con la creación de una inteligencia "artificial".

En su Metafísica, que comienza con las palabras "Todos los hombres desean por naturaleza saber", Aristóteles desarrolló una ciencia de las cosas que nunca cambian, incluyendo su cosmología y teología. Más relevante a la inteligencia artificial, sin embargo, fue la epistemología de Aristóteles, o el análisis de cómo los seres humanos "conocen" su mundo, discutido en su lógica. Aristóteles se refiere a la lógica como el "instrumento" (*Organon*), porque sentía que el estudio del pensamiento mismo se encontraba en la base de todo conocimiento. En su lógica, se investigó si ciertas proposiciones se puede decir que son "verdaderas", ya que se relacionan con otras cosas que son conocidos por ser "verdaderas". Por lo tanto, si sabemos que "todos los hombres son mortales" y que "Sócrates es un hombre", entonces podemos concluir que "Sócrates es mortal". Este argumento es un ejemplo de lo que Aristóteles denomina un silogismo, usando la forma deductiva *modus ponens*. Aunque la axiomatización formal de razonamiento necesita otros dos mil años para su plena floración en los trabajos de GottlobFrege, Bertrand Russell, KurtGödel, Alan Turing, Alfred Tarski, y otros, sus raíces se pueden remontar a Aristóteles.

Pensamiento renacentista, basándose en la tradición griega, inició la evolución de una manera diferente y poderosa de pensar acerca de la humanidad y su relación con el mundo natural. La ciencia empezó a reemplazar el misticismo como medio para entender la naturaleza. Los relojes y, eventualmente, los horarios de fábrica sustituirán a los ritmos de la naturaleza para miles de habitantes de las ciudades. La mayoria de las ciencias sociales y físicas encontraron su origen en la idea de que los procesos, ya sea natural o artificial, podría ser analizado y comprendido matemáticamente. En particular, los científicos y los filósofos se dieron cuenta de que el pensamiento mismo, o la forma en que el conocimiento se ha representado y manipulado en la mente humana, era un tema difícil pero esencial para el estudio científico.

Tal vez el principal evento en el desarrollo de la visión del mundo moderno fue la revolución Copernicana, la sustitución del antiguo modelo geocéntrico del universo con la idea de que la Tierra y otros planetas en realidad están en órbitas alrededor del sol. Después de siglos de una orden "obvia", en el que la explicación científica de la naturaleza del cosmos era coherente con las enseñanzas de la religión y el sentido común, un modelo drásticamente diferente y en absoluto se propone un modelo obvio para explicar los movimientos de los cuerpos celestes. Tal vez por primera vez, *nuestras ideas sobre el mundo eran visto como fundamentalmente distintas de la apariencia.* Esta división entre la mente humana y su realidad en torno, entre las ideas sobre las cosas y las propias cosas, es esencial para elestudio moderno de la mente y su organización. Esta brecha se amplió por los escritos de Galileo, cuyas observaciones científica contradice las verdades más "obvios" sobre el mundo natural y cuyo desarrollo de las matemáticas como una herramienta para describir que mundo hizo hincapié en la distinción entre el mundo y nuestras ideas al respecto. Es a partir de esta infrinjan que la noción moderna de la mente evolucionó: la introspección se convirtió en un motivo común en la literatura, los filósofos comenzaron a estudiar la epistemología y las matemáticas, así como la aplicación sistemática del método científico rivalizaba con los sentidos como herramientas para entender el mundo.

En 1620, *Novum Organun*, de Francis Bacon  ofreció un conjunto de técnicas de búsqueda para esta metodología científica emergente. Sobre la base de la idea Aristotélica y Platónica de que el "forma" de una entidad era equivalente a la suma de sus “características” necesarias y suficientes, Bacon articula un algoritmo para determinar la esencia de una entidad. En primer lugar, se realizó una colección organizada de todas las instancias de la entidad, enumerando las características de cada una en una mesa. Luego se recoge una lista similar de casos negativos de la entidad, centrándose especialmente en instancias próximas de la entidad, es decir, aquellas que se desvió de la "forma" de la entidad por las características individuales. A continuación, Bacon intenta - este paso no es totalmente claro – crear una lista sistemática de todas las características esenciales de la entidad, es decir, aquellas que son comunes a todos los casos positivos de la entidad y que faltan en los casos negativos.

Es interesante ver una forma de enfoque de Francis Bacon a aprendizaje de conceptos reflejado en los modernos algoritmos de inteligencia artificial en “Buscar en espacios de versiones” en Capítulo 10.2. Una extensión de algoritmos de Bacon fue también parte de un programa de IA para el aprendizaje por descubrimiento, adecuadamente llamado *Bacon* (Langley et al. 1981). Este programa fue capaz de inducir muchas leyes físicas a partir de colecciones de datos relacionados con los fenómenos. También es interesante observar que la cuestión de si un algoritmo de uso general era posible para producir pruebas científicas aguardando los desafíos de matemáticas de principios del siglo XX Hilbert (su obra *Entscheidungsproblem)* y la respuesta del genio moderno Alan Turing *(*su *Máquina de Turing*  y pruebas de *la computabilidad* y el *problema de la parada);* ver Davis et al. (1976).

Aunque la primera máquina de calcular, el ábaco, fue creado por los chinos en elsiglo XXVI a.C, más mecanización de los procesos algebraicos esperaba a los conocimientos de los europeos del siglo XVII. En 1614, el matemático escocés John Napier, creo logaritmos, las transformaciones matemáticas que permitieron la multiplicación y el uso de exponentes reducidos a la suma y la multiplicación. Napier también creó sus los llamados “*Ossos de Napier*”, que fueron utilizados para representar los valores de desbordamiento para las operaciones aritméticas. Los dispositivos fueron más tarde utilizados por Wilhelm Schickard (1592-1635), un matemático y sacerdote alemán de Tubinga, quien en 1623 inventó un *Reloj de Cálculo* para la realización de sumas y restas. Esta máquina registró el desbordamiento de sus cálculos por el repique de un reloj.

Otra máquina calculadora famosa fue la *Pascaline* que Blaise Pascal, filósofo y matemático franceses, creo en 1642. Aunque los mecanismos de Schickardy Pascal se limitan a la suma y la resta - incluyendo acarreos y toma prestada – que mostraron que los procesos que anteriormente se pensaba que requieren pensamiento y de habilidad humano podría ser totalmente automatizado. Conforme Pascal cito posteriormente en sus *Pensamientos* (1670), La “la máquina aritmética produce efectos que llegan más a pensar que todas las acciones de los animales”.

Los éxitos de Pascal con máquinas de calcular inspiro a Gottfried Leibniz en 1694 para completar una máquina de trabajo que se conoce como la *Rueda de Leibniz* Se integró un tambor móvil y manivela para impulsar las ruedas y los cilindros que llevan a cabo las operaciones más complejas de multiplicación y división. Leibniz también estaba fascinado por la posibilidad de una lógica automatizada para pruebas de proposiciones. Volviendo al algoritmo de especificación entidad de Bacon, en donde los conceptos se caracterizaron como la colección de sus atributos necesarios y suficientes, Leibniz imagino una máquina capaz de usar esas características para producir conclusiones lógicamente correctas. Leibniz (1887) también idealiza una máquina que refleja las ideas modernas de la inferencia deductiva y la prueba, por lo que la producción de conocimiento científico podría llegar a ser automatizado, un cálculo para el razonamiento.

Los siglos XVII y XVIII vieron también una gran parte de la discusión de las cuestiones epistemológicas; tal vez el más influyente fue la obra de René Descartes, una figura central en el desarrollo de los conceptos modernos de pensamiento y teorías de la mente. En su obra *Meditaciones*, Descartes (1680) trató de encontrar una base para la realidad puramente través de la introspección. Rechazando sistemáticamente la entrada de sus sentidos como poco fiable, Descartes fue forzado a dudar incluso de la existencia del mundo físico y se quedó sólo con la realidad del pensamiento; incluso su propia existencia tenía que ser justificada en términos de pensamiento: "*Cogito ergo sum*" (pienso, luego existo). Después de que él estableció su propia existencia puramente como una entidad pensamiento, Descartes deduce la existencia de Dios como un creador esencial y en última instancia reafirmó la realidad del universo físico como la creación necesaria de un Dios generoso.

Podemos hacer dos observaciones: primero, el división entre la mente y el mundo físico se ha vuelto tan completa que el proceso de pensamiento podría ser discutido al margen de cualquier entrada sensorial o emitir materiales específicos; segundo, la conexión entre la mente y el mundo físico era tan tenue que requiere la intervención de un Dios generoso para mantener el conocimiento fiable del mundo físico! Este punto de vista de la dualidad entre la mente y el mundo físico subyace a todo el pensamiento de Descartes, incluso del desarrollo de la geometría analítica. ¿De qué otra forma podía ser unificada una rama de la matemática aparentemente tan mundana como la geometría con una estructura matemática tan abstracta como el álgebra?

¿Por qué hemos incluido esta discusión mente/cuerpo en un libro sobre la inteligencia artificial? Hay dos consecuencias de este análisis esencial para el entendimiento de la IA:

1. Al tratar de separar la mente del mundo físico, Descartes y otros pensadores relacionados establecieron que la estructura de las ideas sobre el mundo no era necesariamente la misma que la estructura de su objeto. Que está detrás de la metodología de AI, juntamente con los campos de la epistemología, la psicología, la mayor parte de las matemáticas superiores, y la mayoría de la literatura moderna: los procesos mentales tienen una existencia propia, obedecen a sus propias leyes, y pueden ser estudiados por sí mismo.
2. Una vez que la mente y el cuerpo se separan, filósofos vieron la necesidad de encontrar una manera de volver a conectarlos, porque la interacción entre los planos mental, *res cogitans*, y físico, *res extensa*, de Descartes es esencial para la existencia humana.

Aunque millones de palabras han sido escritas sobre el *problema mente-cuerpo*, y diversas soluciones propuestas, nadie ha explicado satisfactoriamente las interacciones obvias entre los estados mentales y las acciones físicas, puede valer una diferencia fundamental entre ellos. La respuesta más aceptada para este problema, y aquella que proporciona una base esencial para el estudio de IA, sostiene que la mente y el cuerpo no son, de ningún modo, entidades fundamentalmente diferentes. Desde esta perspectiva, los procesos mentales son realmente alcanzados por sistemas físicos, tales como el cerebro (u ordenadores). Los procesos mentales, al igual que los procesos físicos, por fin pueden ser caracterizados a través de las matemáticas formales. o bien, como se reconoce en *Leviatán,* por el filósofo Inglés del siglo XVII Thomas Hobbes (1651), "por el razonamiento, quiero decir cálculo".

**1.1.2 IA y la racionalista y tradiciones empiristas**

Cuestiones modernas de investigación en inteligencia artificial, como en otras disciplinas científicas, se forman y evolucionan a través de una combinación de presiones históricas, sociales y culturales. Dos de las presiones más importantes para la evolución de la IA son los tradiciones empiristas y racionalistas de la filosofía.

La tradición racionalista, como vimos en la sección anterior, tenía un autor temprano en Platón, y fue continuada por los escritos de Pascal, Descartes y Leibniz. Para el racionalista, el mundo exterior se reconstruye a través de las ideas claras y distintas de las matemáticas. Una crítica de este enfoque dualista es la separación forzada de los sistemas de representación de su campo de referencia. La cuestión es si el significado que se atribuye a una representación puede ser definido independientemente de sus condiciones de aplicación. Si el mundo es diferente de nuestras creencias sobre el mundo, pueden nuestros conceptos y símbolos creados todavía tener significados?

Muchos programas de IA tienen mucho de este aspecto racionalista. Los primeros planificadores robots, por ejemplo, describirían su dominio de aplicación o el "mundo" como conjuntos de declaraciones de cálculos de predicados, y luego un "plan" para la acción sería creado a través de la demostración de teoremas acerca de este "mundo" (Fikes et al. 1972, véase también sección 8.4). *La física de Símbolos del Sistema Hipótesis* de Newell y Simon (Introducción a la Parte II y en el capítulo 16) es vista por muchos como el arquetipo de este enfoque en la IA moderna. Duras críticas eligen este sesgo racionalista como parte del fracaso de la IA en la resolver tareas complejas, tales como la comprensión de los lenguajes humanos (Searle 1980, Winograd y Flores 1986, Brooks 1991a).

En lugar de afirmar como "real" el mundo de las ideas claras y distintas, empiristas siguen recordándonos que "nada entra en la mente sino a través de los sentidos". Esta restricción lleva a más preguntas de cómo el humano posiblemente puede percibir conceptos generales o las formas puras de la caverna de Platón (Platón, 1961). Aristóteles era un empírico temprano, haciendo hincapié en su *De Anima*, las limitaciones del sistema perceptivo humano. Empiristas más modernos, especialmente Hobbes, Locke y Hume, hacen hincapié en que el conocimiento debe ser explicado a través de una psicología introspectiva pero empírica. Se distinguen dos tipos fenómenos mentales, la percepción, de un lado y el pensamiento, la memoria y la imaginación en el otro.

El filósofo escocés David Hume, por ejemplo, distingue *impresiones* de *ideas*. Las impresiones son dinámicas y vividas, lo que refleja la presencia y la existencia de un objeto externo y no sujeto a control voluntario, los *qualia* de Dennett (2005). Las ideas, por otro lado, son menos vívidas y detalladas y más sujetos a control voluntario del sujeto.

Teniendo en cuenta esta distinción entre impresiones e ideas, ¿cómo puede surgir el conocimiento? Para Hobbes, Locke, Hume y el mecanismo explicativo fundamental es la *asociación*.Determinadas propiedades perceptivas son asociadas por medio de la experiencia repetitiva. Esta asociación repetida crea una disposición en la mente para asociar las ideas correspondientes, un precursor del enfoque conductista del siglo XX. Una propiedad fundamental de esta cuenta se presenta con el escepticismo de Hume. La explicación puramente descriptiva de Hume de los orígenes de las ideas no puede, según él, apoyarse la creencia en la causalidad. Incluso el uso de la lógica y la inducción no puede ser apoyada de manera racional en esta epistemología empirista radical.

En una Investigación sobre el entendimiento humano (1748), el escepticismo de Hume se extiende para el análisis de los milagros. Aunque Hume no se refirió a la naturaleza de los milagros directamente, se puso en duda la creencia milagrosa basada en el testimonio. Este escepticismo, por supuesto, fue visto como una amenaza directa por los creyentes en la Biblia, así como muchos otros proveedores de tradiciones religiosas. El reverendo Thomas Bayes era a un matemático y un ministro religioso. Uno de sus trabajos, llamado *Ensayo hacia la solución de un problema en la Doctrina de Oportunidades* (1763) aborda las cuestiones de Hume matemáticamente. El teorema de Bayes demuestra formalmente cómo podemos, mediante el aprendizaje de las correlaciones de los efectos de las acciones, determinar la probabilidad de sus causas.

La explicación asociativa del conocimiento desempeña un papel significativo en el desenvolvimiento de estructuras y los programas representativos de IA, por ejemplo, en la memoria de la organización con las redes semánticas y MOPS y el trabajo en la comprensión del lenguaje natural (ver secciones 7.0, 7.1, y Capítulo 15). Las explicaciones asociativas tienen influencias importantes del aprendizaje de máquina, especialmente con las redes conexionistas (véase la sección 10.6, 10.7, y en el capítulo 11). Asociacionismo también desempeña un papel importante en la psicología cognitiva incluyendo los *esquemas* de Bartlett y Piaget, así como toda la confianza de la tradición conductista (Luger 1994). Por último, con herramientas de IA para el análisis estocástico, incluyendo la *red Bayesiana* *de creencia* (RBC o BBN) y sus extensiones actuales a los sistemas complejos de Turing de primer orden para la modelización estocástica, las teorías asociativas encontraran una base matemática sólida y una potencia de expresión madura. Las herramientas Bayesianos importantes para la investigación incluyendo diagnósticos, aprendizaje automático, y la comprensión del lenguaje natural (véanse los capítulos 5 y 13).

Emmanuel Kant, filósofo alemán formado en la tradición racionalista, era fuertemente influenciado por los escritos de Hume. Como resultado de ello, comenzó la síntesis moderna de estas dos tradiciones. El conocimiento de Kant contiene dos energías colaboradoras, un componente *a priori* que viene de la razón del sujeto junto con un componente *posteriori* procedente de experiencia activa. La experiencia es significativa sólo a través de la contribución del sujeto. Sin una forma organizadora activa propuesta por el sujeto, el mundo no sería nada más que sensaciones transitorias pasajeras. Por último, a nivel de juicio, las afirmaciones de Kant, imágenes o representaciones que pasan están unidas por el sujeto activo y tomaron como los diversos aspectos de una identidad, de un "objeto". El realismo de Kant comenzó la empresa moderna de los psicólogos como Bartlett, Brunner, y Piaget. La obra de Kant influye en la iniciativa moderna de la IA del aprendizaje automático (sección IV), así como el desarrollo continuo de una epistemología constructivista (véase el Capítulo 16).

**1.1.3 El desarrollo de la lógica formal**

Una vez que el pensamiento comenzó a ser considerado como una forma de cálculo, su formalización y eventual mecanización eran evidentes los siguientes pasos. Como se ha señalado en la sección 1.1.1, Gottfried Leibniz, con su *Calculus Philosophicus*, introdujo el primer sistema de lógica formal, Además de proponer una máquina para automatizar sus tareas (Leibniz 1887). Además, los pasos y las etapas de esta solución mecánica pueden ser representados como movimiento a través de los estados de un árbol o gráfico. Leonhard Euler, en el siglo XVIII, con su análisis de la "conexión" de los puentes que unen las orillas de los ríos y las islas de la ciudad de Königsberg (ver la introducción del capítulo 3), presentó el estudio de las representaciones que se pueden capturar de forma abstracta la estructura de relaciones en el mundo, así como los pasos discretos dentro de un cálculo sobre estas relaciones (Euler, 1735).

La formalización de la teoría de grafos también la posibilidad de la *búsqueda de espacio de estado*, una herramienta conceptual importante de la inteligencia artificial. Podemos usar grafos para modelar la estructura más profunda de un problema. Los nodos de un *grafo de espacio de estado* representan espacio de estados posibles de solución de un problema; los arcos del grafo representan inferencias, se mueve en un juego, u otros pasos en una solución de un problema. La solución del problema es un proceso de búsqueda el grafo de espacio de estados, un camino para una solución (Introducción a la II y Capítulo 3). Mediante la descripción de todo el espacio de soluciones a los problemas, los grafos de espacio de estado proporcionan una herramienta poderosa para la medir la estructura y la complejidad de los problemas y analizar la eficiencia, la exactitud, y la generalidad de estrategias de solución.

Como uno de los creadores de la ciencia de la investigación de operaciones, así como el diseñador de las primeras máquinas programables de cálculo mecánico, Charles Babbage, matemático del siglo XIX, puede también ser considerado como uno de los pioneros de la inteligencia artificial (Morrison y Morrison, 1961). La *máquina diferencial* de Babbage era una máquina para fines especiales para el cálculo de los valores de ciertas funciones polinómicas y fue el precursor de la *máquina analítica*. La máquina analítica, diseñado pero no construido con éxito durante su vida, era una máquina computacional programable de propósito general que presagiaba muchas de las suposiciones arquitectónicas que forman la base del ordenador moderno.

En la descripción de la máquina analítica, Ada Lovelace (1961), amiga incentivadora y colaboradora de Babbage, dice:

Podemos afirmar muy apropiadamente que la máquina analítica teje patrones algebraicos del mismo modo que el telar de Jacquard teje flores y hojas. Aquí, al parecer, se encuentra mucho más originalidad que podría atribuirse a la máquina diferencial.

La inspiración de Babbage fue su deseo de aplicar la tecnología de su tiempo para liberar a los seres humanos de la monotonía de hacer cálculos aritméticos. En este sentimiento, así como con su concepción de las computadoras como dispositivos mecánicos, Babbage estaba pensando en términos del siglo XIX pura. Su motor analítico, sin embargo, también incluye muchas nociones modernas, tales como la separación de la memoria y el procesador, el *depósito* y el *molino* en términos de Babbage, el concepto de una maquina digital en lugar de máquina analógica y capacidad de programación basada en la ejecución de una serie de operaciones codificadas en tarjetas de cartón perforadas. La característica más impresionante la descripción de Ada Lovelace, y en toda la obra de Babbage, en general, es el tratamiento de los "patrones" de las relaciones algebraicas como entidades que pueden ser estudiadas, caracterizadas y finalmente implementados y manipulados mecánicamente sin la preocupación por los valores particulares que son, pasados por el “molino” de la máquina de calcular. Este es un ejemplo de implementación de la "abstracción y la manipulación de la forma" descrita por primera vez por Aristóteles y Leibniz.

El objetivo de crear un lenguaje formal para el pensamiento también aparece en la obra de George Boole, matemático del siglo XIX, cuya obra debe ser incluido en cualquier discusión sobre las raíces de la inteligencia artificial (Boole, 1847, 1854). A pesar de que hizo contribuciones en varias de áreas de las matemáticas, su trabajo más conocido fue en la formalización matemática de las leyes de la lógica, un logro que constituye el corazón de la informática moderna. Aunque el papel del álgebra de Boole en el diseño de circuitos lógicos es bien sabido, los propios objetivos de Boole en el desarrollo de su sistema parecen más a los de los investigadores contemporáneos de IA. En el primer capítulo de *una investigación de las leyes del pensamiento, en que se basan las teorías matemáticas de la lógica y las probabilidades*, Boole (1854) describe sus objetivos como

Investigar las leyes fundamentales de las operaciones mentales por las que uno realiza el razonamiento: las expresó en el lenguaje simbólico de un cálculo y, sobre esa base, establecer la ciencia de la lógica y de dar instrucciones a su método: .. Y finalmente recoger diversos elementos de verdad, revelada en el curso de estas consultas, algunas indicaciones probable relación con la naturaleza y la constitución de la mente humana.

La importancia de la realización de Boole en este extraordinario poder y la simplicidad del sistema que ideó: tres operaciones, "Y" (representado por \* o ∧), "O" (representado por + o ∨), y "NO" (representado por ¬), formó el corazón de su cálculo lógico. Estas operaciones forman la base para todos los desarrollos posteriores de la lógica formal, incluyendo el diseño de los ordenadores modernos. Mientras se mantiene el significado de estos símbolos casi idénticas a las operaciones algebraicas correspondientes, Boole señaló que "los símbolos de la lógica son más sujetos a una ley especial, a la que los símbolos de cantidad, como tales, no están sujetos". Esta ley establece que para cualquier X, que es un elemento del álgebra, X \* X = X (o algo que una vez que se sabe que es verdad, la repetición no puede aumentar ese conocimiento). Esto llevó a la restricción característica de los valores booleanos para los dos únicos números que pueden satisfacer esta ecuación: 1 y 0. Las definiciones estándar de multiplicación booleana (AND) y adición (O) se derivan de esta idea.

El sistema de Boole no sólo proporciona la base de la aritmética binaria, sino también demostró que un sistema formal extremadamente simple era adecuado para capturar toda la potencia de la lógica. Esta suposición y el sistema que Boole desarrolló para demostrar que forman la base de todos los esfuerzos modernos para formalizar la lógica, desde la *Principia Matemática* (Principios Matemáticos) (Whitehead y Russell 1950), a través del trabajo de Turing y Gödel, hasta los sistemas modernos de razonamientos automatizados.

Gottlob Frege, en su obra *Fundamentos de la Aritmética* (Frege, 1879, 1884), creó un lenguaje de especificación matemática para describir la base de la aritmética de una manera clara y precisa. Con este lenguaje Frege formalizado muchos de los temas abordados por primera vez por la *Lógica de Aristóteles*. El lenguaje de Frege, que ahora se llama como *cálculo de predicados de primer orden*, ofrece una herramienta para describir las proposiciones y las asignaciones de valor de verdad que componen los elementos de razonamiento matemático y describe la base axiomática de "significado" de estas expresiones. El sistema formal del cálculo de predicados, que incluye símbolos de predicado, una teoría de funciones y variables cuantificadas, estaba destinado a ser un lenguaje para describir las matemáticas y sus bases filosóficas. También juega un papel fundamental en la creación de una teoría de la representación de la inteligencia artificial (Capítulo 2). el cálculo de predicados de primer orden proporciona herramientas necesarias para el razonamiento automatizado: un lenguaje de expresiones, para unir las hipótesis teóricas relacionadas con el significado de las expresiones y el cálculo correcto para inferir lógicamente nuevas expresiones verdaderas.

Whitehead y Russell (1950) trabajo es particularmente importante para las bases de IA, en la que su objetivo declarado era obtener toda la matemática a través de operaciones formales en una colección de axiomas. Aunque muchos sistemas matemáticos se han construido a partir de axiomas básicos, lo interesante aquí es el comportamiento de Russell y Whitehead con las matemáticas como un sistema puramente formal. Esto significaba que los axiomas y teoremas serían tratados únicamente como cadenas de caracteres: las pruebas se procesarían a través de la aplicación de reglas bien definidas para la manipular estas cadenas. No habría ninguna confianza en la intuición o el significado de teoremas como base para pruebas. Cada paso de una prueba y resultado de la aplicación estricta de las reglas formales (sintácticas) a cualquiera de los axiomas o teoremas previamente probadas, incluso cuando las pruebas tradicionales podrían considerar ese paso como "evidente". Lo, "que significa" los teoremas y axiomas del sistema podrían tener en relación con el mundo sería independiente de sus derivaciones lógicas. Este tratamiento de razonamiento matemático en términos puramente formales (y por lo tanto mecánicos) proporciona una base esencial para su automatización en equipos físicos. Las reglas sintácticas y lógicas formales de inferencia desarrollados por Russell y Whitehead siguen siendo una base para los sistemas automáticos de medida de teorema, que se presentan en el capítulo 14, así como de los fundamentos teóricos de la inteligencia artificial.

Alfred Tarski es otro matemático cuyo trabajo es esencial para las bases deAI. Tarski creó una teoría de la referencia en el que la fórmula bien formadade Frege o Russell y Whitehead se puede decir para referirse, de manera precisa, con el mundo físico (Tarski 1944, 1956; véase el capítulo 2). Esta visión subyace en la mayoría de las teorías de la semántica formal.

En su artículo La concepción semántica de la verdad y la Fundación de la Semántica, Tarski describe su teoría de las relaciones de referencia y el valor de verdad. Los informáticos modernos, especialmente Scott, Strachey, Burstall (Burstall y Darlington 1977), y Plotkin han relacionado con esta teoría de lenguajes de programación y otras especificaciones para la informática.

Aunque en los siglos XVIII, XIX y principios del XX la formalización de la ciencia y las matemáticas creadas el requisito previo intelectual para el estudio de la inteligencia artificial, no fue hasta el siglo XX y la introducción de la computadora digital que AI se convirtió en una disciplina científica viable. A finales de la década de 1940 los ordenadores digitales electrónicos habían demostrado su potencial para proporcionar la capacidad de memoria y procesamiento requerido por los programas inteligentes. Ahora era posible implementar sistemas formales de razonamiento en un equipo y empíricamente a prueba su suficiencia para la exposición de la inteligencia.

Un componente esencial de la ciencia de la inteligencia artificial es este compromiso a las computadoras digitales como el vehículo de elección para crear y probar las teorías de la inteligencia.

Las computadoras digitales no son simplemente un vehículo para probar teorías de la inteligencia. Su arquitectura también sugiere un paradigma específico para tales teorías: inteligencia es una forma de tratamiento de la información. La noción de búsqueda como una metodología de resolución de problemas, por ejemplo, se debe más a la naturaleza secuencial de funcionamiento del ordenador que lo hace para cualquier modelo biológico de la inteligencia.

La mayoría de los programas de IA representan el conocimiento en un lenguaje formal que a continuación se gestiona mediante algoritmos, en honor a la separación de los datos y el programa fundamental para el estilo de von Neumann de la computación. La lógica formal se ha convertido en una herramienta de representación importante para la investigación en IA, al igual que la teoría de grafos juega un papel indispensable en el análisis de los espacios de problemas, así como proporcionar una base para las redes semánticas y modelos similares de significado semántico. Estas técnicas y formalismos se discuten en detalle en todo el cuerpo de este texto; Los mencionamos aquí para enfatizar la relación simbiótica entre el ordenador digital y los fundamentos teóricos de la inteligencia artificial.

**1.1.4El Test de Turing**

Uno de los primeros trabajos para abordar la cuestión de la inteligencia de las máquinasespecíficamente en relación con la computadora digital moderna fue escrito en 1950 por el matemático británico Alan Turing. Maquinaria de Computación e Inteligencia (Turing 1950) sigue siendo oportuna, tanto en su evaluación de los argumentos contra la posibilidad de crear una máquina de computación inteligente y sus respuestas a esos argumentos. Turing, conocido principalmente por sus contribuciones a la teoría de la computabilidad, examinó la cuestión de si es o no una máquina en realidad se podría hacer pensar. Tomando nota de que las ambigüedades fundamentales en la pregunta misma (lo que está pensando? ¿Qué es una máquina?) Excluyeron cualquier respuesta racional, propuso que la cuestión de la inteligencia ser reemplazada por una prueba empírica más claramente definido.

Las medidas de prueba de Turing el rendimiento de una máquina supuestamente inteligente frente a la de un ser humano, sin duda el mejor y único estándar para un comportamiento inteligente. La prueba, que Turing llamado el juego de imitación, coloca la máquina y un homólogo humano en habitaciones aparte de un segundo ser humano, denominado como el interrogador (Figura 1.1). El interrogador no es capaz de ver o hablar directamente a cualquiera de ellos, no sabe qué entidad es en realidad la máquina, y puede comunicarse con ellos únicamente mediante el uso de un dispositivo textual tal como un terminal. Se le pide al interrogador para distinguir el equipo desde el ser humano únicamente sobre la base de sus respuestas a las preguntas formuladas sobre este dispositivo. Si el interrogador no puede distinguir la máquina de la humana, entonces, Turing argumenta, la máquina puede suponerse que es inteligente.

Al aislar el interrogador de la máquina y el otro participante humano, la prueba se asegura de que el interrogador no se hará con preferencia por la aparición de la máquina o de cualquier propiedad mecánica de su voz. El interrogador es libre, sin embargo, para hacer cualquier pregunta, no importa qué tan tortuosa o indirecta, en un esfuerzo por descubrir la identidad del equipo. Por ejemplo, el interrogador puede preguntar a ambos sujetos para realizar un cálculo aritmético bastante complejo, en el supuesto de que el equipo será más probable que lo correcto que el ser humano; para contrarrestar esta estrategia, el equipo tendrá que saber cuándo debe dejar de obtener una respuesta correcta a este tipo de problemas con el fin de parecer como un ser humano. Para descubrir la identidad del ser humano sobre la base de la naturaleza emocional, el interrogador puede preguntar a los dos sujetos para responder a un poema o una obra de arte; Esta estrategia requerirá que el ordenador tenga conocimiento de la estructura emocional de los seres humanos.



Las características importantes de la prueba de Turing son:

1. Se intenta dar una idea objetiva de la inteligencia, es decir, el comportamiento de un ser inteligente conocido en respuesta a un determinado conjunto de preguntas. Esto proporciona un estándar para la determinación de la inteligencia que evita los debates inevitables sobre su naturaleza "verdadera".

2. Nos impide ser desviado por tales preguntas confusas y actualmente incontestables como si el equipo utiliza los procesos internos adecuados o si o no la máquina es realmente consciente de sus acciones.

3. Se elimina cualquier sesgo a favor de los organismos vivos al forzar el interrogador para centrarse únicamente en el contenido de las respuestas a las preguntas.

Debido a estas ventajas, la prueba de Turing proporciona una base para muchos de los esquemas utilizados efectivamente para evaluar los programas de IA modernas. Un programa que ha logrado potencialmente inteligencia en alguna área de especialización se puede evaluar mediante la comparación de su desempeño en un determinado conjunto de problemas a la de un experto humano. Esta técnica de evaluación es sólo una variación de la prueba de Turing: Se recuerda a un grupo de humanos para comparar ciegamente el rendimiento de un ordenador y un ser humano en un conjunto particular de problemas. Como veremos, esta metodología se ha convertido en una herramienta esencial en el desarrollo y la verificación de los sistemas expertos modernos.

El test de Turing, a pesar de su atractivo intuitivo, es vulnerable a una serie de críticas justificadas. Uno de los más importantes de ellos se dirige a su sesgo hacia las tareas puramente simbólicas de resolución de problemas. No prueba las habilidades que requieren habilidades de percepción o de destreza manual, a pesar de que estos son componentes importantes de la inteligencia humana. Por el contrario, a veces se sugiere que la prueba de Turing limita innecesariamente la inteligencia artificial para adaptarse a un molde humano. Tal vez la inteligencia artificial es simplemente diferente de la inteligencia humana y tratar de evaluar en términos humanos es un error fundamental.

Es lo que realmente deseamos una máquina haríamos las matemáticas tan lenta y erróneamente como un ser humano? ¿No debería una máquina inteligente sacar partido de sus propios activos, tales como una memoria de gran tamaño, rápido, fiable, en lugar de tratar de emular la cognición humana? De hecho, un número de practicantes de AI modernos (por ejemplo, Ford y Hayes 1995) ver en respuesta a la plena reto de la prueba de Turing como un error y una gran distracción para el trabajo más importante que nos ocupa: el desarrollo de teorías generales para explicar los mecanismos de la inteligencia en los seres humanos y máquinas para la aplicación de esas teorías para el desarrollo de herramientas para resolver problemas específicos y prácticos. Aunque estamos de acuerdo con las preocupaciones de Ford y Hayes en el intestino, todavía vemos la prueba de Turing como un componente importante en la verificación y validación de software moderno AI.

Turing también se dirigió a la propia viabilidad de construir un programa inteligente en un ordenador digital. Al pensar en términos de un modelo específico de cálculo (una máquina de computación de estado discreto electrónica), hizo algunas conjeturas bien fundadas acerca de la capacidad de almacenamiento, la complejidad del programa, y ​​la filosofía de diseño básico requerido para un sistema de este tipo. Por último, se dirigió a una serie de objeciones morales, filosóficas y científicas a la posibilidad de construir un programa de este tipo en términos de una verdadera tecnología. Se remite al lector al artículo de Turing para un resumen perceptiva y siempre actual del debate sobre la posibilidad de que las máquinas inteligentes.

Dos de las objeciones citadas por Turing son dignos de considerarse más. Objeción de Lady Lovelace, declarado por primera vez por Ada Lovelace, sostiene que los ordenadores sólo pueden hacer lo que se les dice y por lo tanto no pueden realizar acciones originales (por lo tanto, inteligentes). Esta objeción se ha convertido en una parte algo dudoso si tranquilizador del folklore tecnológico contemporáneo. Los sistemas expertos (Sección 1.2.3 y el capítulo 8), especialmente en el área de razonamiento diagnóstico, han llegado a conclusiones no anticipado por sus diseñadores. De hecho, un número de investigadores creen que la creatividad humana se puede expresar en un programa de ordenador.

La otra objeción relacionada, el argumento de la informalidad de la acción, afirma la imposibilidad de crear un conjunto de reglas que le dirá a un individuo exactamente qué hacer bajo cualquier posible conjunto de circunstancias. Ciertamente, la flexibilidad que permite una inteligencia biológica para responder a una gama casi infinita de situaciones de manera razonable si la moda no necesariamente óptima es una característica de la conducta inteligente. Si bien es cierto que la estructura de control utilizado en la mayoría de los programas de ordenador tradicionales no demuestra una gran flexibilidad o la originalidad, no es cierto que todos los programas deban ser escritos de esta manera.

De hecho, gran parte del trabajo de AI en los últimos 25 años ha sido el desarrollo de lenguajes de programación y modelos tales como los sistemas de producción, sistemas basados ​​en objetos, representaciones de redes neuronales, y otros tratados en este libro que tratan de superar esta deficiencia.

Muchos programas de IA modernos consisten en una colección de componentes modulares, o reglas de comportamiento, que no se ejecutan en un orden rígido, sino más bien se invocan como sea necesario en respuesta a la estructura de un ejemplo de un problema en particular. Comparadores de patrones permiten a las reglas generales que se aplican a través de una serie de casos. Estos sistemas tienen una extrema flexibilidad que permite relativamente pequeños programas para exhibir una amplia gama de posibles comportamientos en respuesta a diferentes problemas y situaciones.

Si estos sistemas en última instancia, se pueden hacer para exhibir la flexibilidad mostrada por un organismo vivo es todavía objeto de mucho debate. El premio Nobel Herbert Simon ha argumentado que gran parte de la originalidad y la variabilidad del comportamiento mostrado por los seres vivos es debido a la riqueza de su entorno en lugar de la complejidad de sus propios programas internos. En las ciencias de lo artificial, Simon (1981) describe una hormiga avanza tortuosamente a lo largo de un tramo irregular y desordenada del suelo. A pesar de que el camino de la hormiga parece bastante complejo, Simon argumenta que el objetivo de la hormiga es muy simple: para volver a su colonia lo más rápidamente posible. Los giros y vueltas en su camino son causados por los obstáculos que encuentra en su camino. Simon llega a la conclusión de que

Una hormiga, visto como un sistema de comportarse, es bastante simple. La aparente complejidad de su comportamiento en el tiempo es en gran parte un reflejo de la complejidad del entorno en el que se encuentra.

Esta idea, si finalmente resultó aplicarse a los organismos de inteligencia superior, así como a tales criaturas simples como insectos, constituye un potente argumento de que tales sistemas son relativamente simples y, por consiguiente, comprensible. Es interesante observar que si se aplica esta idea a los seres humanos, se convierte en un fuerte argumento a favor de la importancia de la cultura en la formación de la inteligencia. En lugar de crecer en la oscuridad como las setas, la inteligencia parece depender de una interacción con un entorno suficientemente rico. La cultura es tan importante en la creación de los seres humanos como seres humanos están en la creación de la cultura. En lugar de denigrar a nuestro intelecto, esta idea hace hincapié en la riqueza milagrosa y la coherencia de las culturas que se han formado fuera de la vida de los seres humanos separados. De hecho, la idea de que la inteligencia surge de las interacciones de los elementos individuales de una sociedad es una de las ideas que apoyan el enfoque a la tecnología de AI se presenta en la siguiente sección.

**1.1.5Modelos Biológicos y Sociales de Inteligencia: Agentes Teorías**

Hasta ahora, hemos abordado el problema de la construcción de máquinas inteligentes desde el punto de vista de las matemáticas, con la creencia implícita de razonamiento lógico como paradigmático de la inteligencia misma, así como con el compromiso de bases "objetivas" para el razonamiento lógico. Esta manera de considerar el conocimiento, el lenguaje y el pensamiento refleja la tradición racionalista de la filosofía occidental, mientras que se desarrolló a través de Platón, Galileo, Descartes, Leibniz, y muchos de los otros filósofos discutidos anteriormente en este capítulo. También refleja las suposiciones subyacentes de la prueba de Turing, en particular su énfasis en el razonamiento simbólico como una prueba de la inteligencia, y la creencia de que una comparación directa con el comportamiento humano era adecuada para confirmar la inteligencia artificial.

La dependencia de la lógica como una forma de representar el conocimiento y en la inferencia lógica como mecanismo principal para el razonamiento inteligente son tan dominante en la filosofía occidental que su "verdad" a menudo parece obvia e inatacable. No es de extrañar, entonces, que los enfoques basados ​​en estos supuestos han dominado la ciencia de la inteligencia artificial a partir de su creación casi hasta el día de hoy.

La segunda mitad del siglo XX ha sido, sin embargo, visto numerosos retos a la filosofía racionalista. Varias formas de relativismo filosófico cuestionan el objetivo base de la lengua, la ciencia, la sociedad y el pensamiento mismo. Más tarde la filosofía de Ludwig Wittgenstein (Wittgenstein 1953), nos ha obligado a reconsiderar la base sobre la que significa en ambos idiomas naturales y formales. El trabajo de Godel (Nagel y Newman 1958) y de Turing ha puesto en duda los fundamentos de la matemática misma. Pensamiento post-moderno ha cambiado nuestra comprensión del significado y el valor de las artes y la sociedad. La inteligencia artificial no ha sido inmune a estas críticas; de hecho, las dificultades que AI ha tropezado en la consecución de sus objetivos a menudo se toman como evidencia del fracaso del punto de vista racionalista (Winograd y Flores 1986, Lakoff y Johnson 1999, Dennett 2005).

Dos tradiciones filosóficas, la de Wittgenstein (1953), así como la de Husserl (1970, 1972) y Heidegger (1962) son fundamentales para la revalorización de la tradición filosófica occidental. En su obra posterior, Wittgenstein cuestionó muchas de las suposiciones de la tradición racionalista, incluyendo los fundamentos del lenguaje, la ciencia y el conocimiento.

El lenguaje humano era un foco importante de análisis de Wittgenstein: desafió la idea de que el lenguaje deriva su significado de cualquier tipo de fundamento objetivo.

Para Wittgenstein, así como la teoría de los actos del habla desarrollada por Austin (1962) y sus seguidores (Grice 1975, Searle 1969), el significado de cualquier enunciado depende de que esté situado en un ser humano, el contexto cultural. Nuestra comprensión del significado de la palabra "silla", por ejemplo, depende de tener un cuerpo físico que se ajusta a una posición de sentado y de las convenciones culturales para el uso de sillas. Cuando, por ejemplo, es una roca grande y plana una silla? ¿Por qué es extraña para referirse al trono de Inglaterra como una silla? ¿Cuál es la diferencia entre la comprensión de un ser humano de una silla y la de un perro o un gato, incapaz de sentarse en el sentido humano? Sobre la base de sus ataques a las bases de sentido, Wittgenstein sostenía que deberíamos ver el uso del lenguaje en términos de decisiones tomadas y las acciones tomadas en un contexto cultural cambiante. Wittgenstein se extendió incluso sus críticas a la ciencia y las matemáticas, argumentando que son apenas tanto construcciones sociales como es el uso del lenguaje.

Husserl (1970, 1972), el padre de la fenomenología, se ha comprometido a abstracciones como enraizadas en el Lebenswelt hormigón o mundo de la vida: un modelo racionalista era mucho secundaria al mundo concreto que lo sustentaba. Para Husserl, así como para su alumno de Heidegger (1962), y su proponente Merleau-Ponty (1962), la inteligencia no era saber lo que era verdadero, sino más bien saber cómo hacer frente a un mundo que estaba en constante cambio y evolución. Gadamer (1976) también contribuyó a esta tradición. Para el existencialista / fenomenólogo, la inteligencia es vista como la supervivencia en el mundo, más que como un conjunto de proposiciones lógicas sobre el mundo (en combinación con algún esquema de inferencia).

Muchos autores, por ejemplo Dreyfus y Dreyfus (1985) y Winograd y Flores(1986), han dibujado en Wittgenstein y la obra de Husserl / Heidegger en sus críticas a la IA. A pesar de que muchos practicantes de AI continuar desarrollando el programa racional / lógica, también conocida como la BAIA, o bien pasado de moda de la IA, un número creciente de investigadores en este campo han incorporado estas críticas en nuevos y emocionantes modelos de inteligencia. De acuerdo con el énfasis de Wittgenstein sobre las raíces antropológicas y culturales del conocimiento, que han convertido a lo social, a veces referido como agente basados ​​en o que se hallen, modelos de comportamiento inteligente para su inspiración.

Como ejemplo de una alternativa a un enfoque basado en la lógica, la investigación en el aprendizaje conexionista (Sección 1.2.9 y el Capítulo 11) resta importancia a la lógica y el funcionamiento de la mente racional, en un esfuerzo para lograr la inteligencia mediante el modelado de la arquitectura de la física cerebro. Modelos neuronales de inteligencia hacen hincapié en la capacidad del cerebro para adaptarse a un mundo en el que se encuentra mediante la modificación de las relaciones entre las neuronas individuales.

En lugar de representar el conocimiento en oraciones lógicas explícitas, lo capturan implícitamente, como una propiedad de los patrones de relaciones.

Otro modelo de base biológica de la inteligencia se inspira en los procesos por los cuales las especies enteras se adaptan a su entorno. El trabajo en la vida artificial y algoritmos genéticos (Capítulo 12) aplica los principios de la evolución biológica de los problemas de encontrar soluciones a problemas difíciles.

Estos programas no resuelven los problemas mediante un razonamiento lógicamente acerca de ellos; más bien, se reproducen las poblaciones de competir soluciones candidatas y los llevan a evolucionar cada vez mejores soluciones a través de un proceso de modelado después de la evolución biológica: soluciones pobres candidatos tienden a desaparecer, mientras que los que muestran la promesa para resolver un problema sobreviven y se reproducen mediante la construcción de nuevas soluciones a partir de componentes de sus padres exitosos.

Los sistemas sociales ofrecen otra metáfora de la inteligencia en el que exhiben comportamientos globales que les permitan resolver los problemas que confundir a cualquiera de sus miembros individuales. Por ejemplo, aunque ningún individuo con precisión podría predecir el número de hogazas de pan que se consume en la ciudad de Nueva York en un día determinado, todo el sistema de las panaderías de Nueva York hace un excelente trabajo de mantener la ciudad abastecido con pan, y hacerlo con un mínimo de desperdicio. El mercado de valores hace un excelente trabajo de establecer los valores relativos de cientos de empresas, a pesar de que cada inversor individual sólo tiene un conocimiento limitado de unas pocas empresas. Un último ejemplo viene de la ciencia moderna. Las personas ubicadas en las universidades, la industria, o entornos de gobierno se centran en los problemas comunes.

Con conferencias y revistas como los principales medios de comunicación, problemas importantes para la sociedad en general son atacados y resueltos por los agentes individuales de trabajo semi-independiente, a pesar de los avances en muchos casos también es impulsado por los organismos de financiación.

Estos ejemplos comparten dos temas: en primer lugar, la vista de la inteligencia, tal como figuraba en la cultura y la sociedad y, como consecuencia, emergente. El segundo tema es que la inteligencia se refleja en los comportamientos colectivos de un gran número de muy simple interactuando, individuos semiautónomas, o agentes. Si estos agentes son células neuronales, los miembros individuales de una especie, o una sola persona en una sociedad, sus interacciones producen inteligencia.

¿Cuáles son los temas principales que apoyan una vista de agente-orientado y emergente de la inteligencia? Incluyen:

1. Los agentes son autónomas o semi-autónoma. Es decir, cada agente tiene ciertas responsabilidades en la resolución de problemas con poco o ningún conocimiento de lo que cualquiera otra agentes hacen o cómo lo hacen. Cada agente hace su propia pieza independiente de la resolución de problemas y, o bien produce un resultado en sí (hace algo) o los resultados de los informes de nuevo a otros en el (agente de comunicación) de la comunidad.

2. Los agentes se "encuentran". Cada agente es sensible a su propio entorno que las rodea y (generalmente) no tiene conocimiento del pleno dominio de todos los agentes. Por lo tanto, el conocimiento de un agente se limita a las tareas a mano: "el archivo-estoy-procesamiento" o "la pared próxima a mí" sin conocimiento de la gama total de archivos o limitaciones físicas en el tarea de resolución de problemas.

3. Los agentes son interaccional. Es decir, que forman un conjunto de individuos que cooperan en una tarea particular. En este sentido, pueden ser vistos como una "sociedad" y, como en la sociedad humana, el conocimiento, las habilidades y responsabilidades, incluso cuando se ve como colectiva, se distribuyen en toda la población de individuos.

4. La sociedad de los agentes está estructurado. En la mayoría de los puntos de vista de la resolución de problemas orientado a agentes, cada individuo, a pesar de tener su propio entorno único y conjunto de habilidades, se coordinará con otros agentes en la solución global problema. Por lo tanto, una solución final será no sólo ser visto como colectiva, sino también como cooperativa.

5. Por último, el fenómeno de la inteligencia en este entorno es "emergente".Aunque los agentes individuales son vistos como poseedores de conjuntos de habilidades y responsabilidades, el resultado global de cooperación puede ser visto como mayor que la suma de sus colaboradores individuales. La inteligencia es vista como un residente en el fenómeno y que emerge de una sociedad y no sólo una propiedad de un agente individual.

Sobre la base de estas observaciones, definimos un agente como un elemento de una sociedad que puede percibir (a menudo limitado) aspectos de su medio ambiente y afectar a ese entorno, ya sea directamente oa través de la cooperación con otros agentes. La mayoría de soluciones inteligentes requieren una variedad de agentes. Estos incluyen agentes de memoria, que simplemente capturar y comunicar piezas de agentes de información, coordinación que pueden apoyar las interacciones entre otros agentes, agentes de búsqueda que se pueden examinar varios fragmentos de información y devolver algún bit elegido de la misma, los agentes del aprendizaje que puede examinar colecciones de información y formar conceptos o generalizaciones, y agentes de decisión que pueden ambas tareas de despacho y llegar a conclusiones a la luz de la información y procesamiento limitada. Volviendo a la definición anterior de la inteligencia, los agentes pueden ser vistos como los mecanismos de apoyo a la toma de decisiones en el contexto de limitados recursos de procesamiento.

Los principales requisitos para el diseño y la construcción de una sociedad como son:

1. Estructuras para la representación de la información,

2. Estrategias para la búsqueda de soluciones alternativas, a través y

3. la creación de arquitecturas que pueden apoyar la interacción de los agentes.

El resto de capítulos de nuestro libro, especialmente la Sección 7.4, incluyen recetas para la construcción de herramientas de apoyo para esta sociedad de agentes, así como muchos ejemplos de resolución de problemas basado en agentes.

Nuestro análisis preliminar de la posibilidad de una teoría de la inteligencia automatizada es en modo alguno destinado a exagerar los progresos realizados hasta la fecha o reducir al mínimo el trabajo que queda por delante. Como destacamos en este libro, es importante ser conscientes de nuestras limitaciones y para ser honestos acerca de nuestros éxitos. Por ejemplo, se han producido escasos resultados con los programas que en ningún sentido interesante se puede decir que "aprender". Nuestros logros para modelar la complejidad semántica de un lenguaje natural tales como inglés también han sido muy modestos. Incluso cuestiones fundamentales tales como la organización del conocimiento o totalmente la gestión de la complejidad y la corrección de los programas de ordenador muy grandes (tales como grandes bases de conocimiento) requieren una considerable investigación adicional. Los sistemas basados ​​en el conocimiento, a pesar de que han logrado éxitos de ingeniería negociables, todavía tienen muchas limitaciones en la calidad y la generalidad de su razonamiento. Estos incluyen su incapacidad para llevar a cabo el razonamiento de sentido común o para exhibir el conocimiento de la realidad física rudimentaria, por ejemplo, cómo las cosas cambian con el tiempo.

Pero debemos mantener una perspectiva razonable. Es fácil pasar por alto los logros de la inteligencia artificial cuando se enfrenta honestamente el trabajo que queda. En la siguiente sección, establecemos esta perspectiva a través de una visión general de varias áreas importantes de la investigación de la inteligencia artificial y el desarrollo.

**1.2Descripción general de áreas de aplicación de IA**

Descripción general de áreas de aplicación de IA

La máquina analítica no tiene pretensiones lo que sea que se originan nada. Se puede hacer lo que sabemos cómo ordenar que éste realice.

-ADA Byron, condesa de Lovelace

*Lo siento, Dave; No puedo dejar que hagas eso.*

Hal 9000 en 2001: Odisea en el espacio de Arthur C. Clarke

Ahora regresamos a nuestro objetivo de definir la inteligencia artificial a través de un examen de las ambiciones y los logros de los trabajadores en el campo. Las dos preocupaciones más fundamentales de los investigadores de IA son la representación del conocimiento y de búsqueda. La primera de estas direcciones el problema de la captura en un idioma, es decir, una adecuada para la manipulación del ordenador, el conjunto de los conocimientos necesarios para un comportamiento inteligente. El capítulo 2 presenta el cálculo de predicados como un lenguaje para describir las propiedades y relaciones entre objetos en dominios de problemas que requieren un razonamiento cualitativo en lugar de cálculos aritméticos para sus soluciones. Más tarde, la Sección III se analiza las herramientas que la inteligencia artificial ha desarrollado para representar las ambigüedades y complejidades de áreas tales como el razonamiento de sentido común y la comprensión del lenguaje natural.

La búsqueda es una técnica de resolución de problemas que explora sistemáticamente un espacio de estados problemáticos, es decir, las etapas sucesivas y alternativa en el proceso de resolución de problemas. Los ejemplos de estados problemáticos podrían incluir las diferentes configuraciones de mesa en un juego o pasos intermedios en un proceso de razonamiento. Este espacio de soluciones alternativas a continuación se realizó búsquedas para encontrar una respuesta. Newell y Simon (1976) han argumentado que esta es la base esencial de la solución de problemas humanos. De hecho, cuando un jugador de ajedrez examina los efectos de los diferentes movimientos o un médico considera que una serie de diagnósticos alternativos, que están buscando entre varias alternativas. Las implicaciones de este modelo y las técnicas para su implementación se discuten en los capítulos 3, 4, 6 y 16. El material auxiliar de este libro (véase el Prefacio) ofrece implementaciones de Lisp, Prolog, y Java de estos algoritmos.

Como la mayoría de las ciencias, la IA se descompone en una serie de sub-disciplinas que, mientras que comparte un enfoque esencial para la resolución de problemas, se han ocupado diferentes aplicaciones. En esta sección describimos varios de estas grandes áreas de aplicación y sus contribuciones a la inteligencia artificial en su conjunto.

.

**1.2.1PlayingGame**

Gran parte de las primeras investigaciones en busca de espacio de estado se realizó utilizando juegos de mesa comunes, como damas, ajedrez, y el 15-rompecabezas. Además de su atractivo intelectual inherente, juegos de mesa tienen ciertas propiedades que los hacen sujetos ideales para la investigación. La mayoría de los juegos se juegan con un conjunto bien definido de reglas: esto hace que sea fácil de generar el espacio de búsqueda y libera el investigador de muchas de las ambigüedades y complejidades inherentes a los problemas menos estructurados. Las configuraciones de cartón usados ​​en jugar juegos son captar las sutilezas semánticas de los dominios de problemas más complejos. Dado que los juegos se juegan fácilmente, poniendo a prueba un programa de juego-juego no presenta ninguna carga financiera o ético.

la búsqueda de espacio de estado, el paradigma que subyace en la mayoría de investigaciones juego de roles, se presenta en los capítulos 3 y 4.

Los juegos pueden generar extremadamente grandes espacios de búsqueda. Estos son grandes y suficientecomplejos como para requerir técnicas de gran alcance para determinar qué alternativas para explorar en el espacio del problema. Estas técnicas se llaman heurísticas y constituyen un área importante de la investigación en IA. Una heurística es una estrategia útil pero potencialmente falibles resolución de problemas, tales como la comprobación para asegurarse de que un aparato no responde está enchufado antes de asumir que es roto o al castillo con el fin de tratar de proteger a su rey, desde la captura de un juego de ajedrez.

Gran parte de lo que comúnmente llamamos inteligencia parece residir en la heurística usada por los seres humanos para resolver problemas.

Dado que la mayoría de nosotros tenemos alguna experiencia con estos juegos simples, es posible idear y poner a prueba la eficacia de nuestra heurística. No necesitamos para encontrar y consultar a un experto en algún área del problema esotérico como la medicina o las matemáticas (el ajedrez es una excepción obvia a esta regla). Por estas razones, juegos ofrecen un dominio rico para el estudio de búsqueda heurística. Capítulo 4 introduce heurística utilizando estos juegos simples.

Programas Juego de papeles, a pesar de su simplicidad, ofrecen sus propios desafíos, entre ellos a un oponente que se mueve no se puede anticipar de manera determinista, capítulos 5 y 9, y la necesidad de considerar los factores psicológicos, así como en tácticas de juego de estrategia.

Los recientes éxitos en juego basado en juego de ordenador incluyen campeonato mundial de backgammon y ajedrez. También es interesante observar que en 2007 el espacio de estados para el juego de damas se trazó, permitiendo que sea desde el primer movimiento, determinista

**1.2.2Razonamiento automatizado y la demostración de teoremas**

Podríamos argumentar que demostración de teoremas automática es la rama más antigua de la inteligencia artificial, remonta sus raíces a través de Newell y Lógico Teórico de Simon (Newell y Simon 1963a) y el General ProblemSolver (Newell y Simon 1963b), a través de los esfuerzos de Russell y Whitehead para tratar todos de las matemáticas como la derivación puramente formal de teoremas a partir de axiomas básicos, a sus orígenes en los escritos de Babbage y Leibniz. En cualquier caso, ha sido sin duda una de las ramas más fructíferas del campo. Teorema de la investigación-probando era responsable de gran parte de los primeros trabajos en la formalización de los algoritmos de búsqueda y el desarrollo de los lenguajes formales de representación, tales como el cálculo de predicados (Capítulo 2) y el lenguaje de programación lógica Prolog.

La mayor parte del atractivo de la demostración automática de teoremas radica en el rigor y la generalidad de la lógica. Debido a que es un sistema formal, la lógica se presta a la automatización. Una amplia variedad de problemas puede ser atacado mediante la representación de la descripción del problema y la información de antecedentes pertinente como axiomas lógicos y el tratamiento de casos de problemas como los teoremas que se probaron. Esta visión es la base del trabajo en la demostración de teoremas automático y sistemas de razonamiento matemático (Capítulo 14).

Por desgracia, los primeros esfuerzos en la escritura de teoremas no desarrollaron un sistema que podría resolver constantemente problemas complicados. Esto era debido a la capacidad de cualquier sistema lógico razonablemente complejo para generar un número infinito de teoremas demostrables: sin técnicas de gran alcance (heurística) para guiar su busca, demostradores automáticos demostraron un gran número de teoremas irrelevantes antes de tropezar con la correcta. En respuesta a esta ineficiencia, muchos argumentan que los métodos puramente formales, sintácticos de búsqueda de guía son inherentemente incapaces de manejar un gran espacio de este tipo y que la única alternativa es confiar en lo informal, estrategias ad hoc que los seres humanos parecen utilizar en la solución de problemas.

Este es el enfoque que subyace en el desarrollo de sistemas expertos (capítulo 8), y ha demostrado ser fructífera.

Aun así, el atractivo de razonamiento basado en la lógica matemática formal es demasiado fuerte como para ignorarlo. Muchos de los problemas importantes, tales como el diseño y verificación de circuitos lógicos, la verificación de la exactitud de los programas de ordenador, y el control de sistemas complejos parecen responder a un enfoque de este tipo. Además, la comunidad teorema-proving ha disfrutado de éxito en la elaboración de potentes heurística de soluciones que se basan únicamente en una evaluación de la forma sintáctica de una expresión lógica, y como resultado, la reducción de la complejidad del espacio de búsqueda sin tener que recurrir a las técnicas especiales utilizado por la mayoría de los solucionadores de problemas humanos.

Otra razón para el continuo interés en teoremas automáticas es la realización de que un sistema de este tipo no tiene que ser capaz de resolver de forma independiente problemas extremadamente complejos sin ayuda humana. Muchos de teoremas modernos funcionan como asistentes inteligentes, los seres humanos dejar que realizan las tareas más exigentes de la descomposición de un gran problema en subproblemas y la elaboración de la heurística para buscar en el espacio de posibles pruebas. El demostrador de teoremas continuación, realiza la más simple pero todavía exigiendo tarea de probar lemas, la verificación de conjeturas más pequeños, y completando los aspectos formales de una prueba esbozados por su asociado humano (Boyer y Moore 1979, Bundy 1988 Veroff 1997 Veroff y Spinks 2006).

**1.2.3Sistemas expertos**

Una visión importante obtenida de los primeros trabajos en la resolución de problemas era la importancia del conocimiento de dominio específico. Un médico, por ejemplo, no es eficaz en el diagnóstico de la enfermedad únicamente porque ella posee cierta habilidad para resolver problemas generales innatos; que es eficaz porque sabe mucho acerca de la medicina. Del mismo modo, un geólogo es eficaz en el descubrimiento de depósitos minerales debido a que es capaz de aplicar una gran cantidad de conocimientos teóricos y empíricos sobre la geología para el problema en cuestión. El conocimiento experto es una combinación de una comprensión teórica del problema y un conjunto de reglas heurísticas de resolución de problemas que la experiencia ha demostrado ser eficaz en el dominio. Los sistemas expertos se construyen mediante la obtención de este conocimiento de un experto humano y la codificación en una forma que un ordenador puede aplicarse a problemas similares.

Esta dependencia de los conocimientos de un experto del dominio humano de estrategias para resolver problemas del sistema es una característica importante de los sistemas expertos. Aunque algunos programas se escriben en el que el diseñador es también la fuente del conocimiento del dominio, es mucho más típico ver a este tipo de programas que crecen fuera de una colaboración entre un experto en el campo, como un médico, químico, geólogo o ingeniero y una por separado especialista en inteligencia artificial. El experto de dominio proporciona los conocimientos necesarios sobre el dominio del problema a través de una discusión general de sus métodos de resolución de problemas y mediante la demostración de estas habilidades en un conjunto cuidadosamente elegida de problemas de muestra. El especialista de la IA, o ingeniero del conocimiento, ya que los diseñadores de sistemas expertos son a menudo conocidos, se encarga de aplicar este conocimiento en un programa que es a la vez eficaz y aparentemente inteligente en su comportamiento. Una vez que un programa de este tipo ha sido escrito, es necesario perfeccionar sus conocimientos a través de un proceso de darle problemas de ejemplo para resolver, dejando que el dominio de expertos critican su comportamiento, y hacer los cambios necesarios o modificaciones en el conocimiento del programa. Este proceso se repite hasta que el programa ha alcanzado el nivel deseado de rendimiento.

Uno de los primeros sistemas para explotar el conocimiento específico del dominio en la resolución de problemas se DENDRAL, desarrollado en Stanford a finales de 1960 (Lindsay et al., 1980). DENDRAL fue diseñado para inferir la estructura de las moléculas orgánicas de sus fórmulas químicas e información espectrográfico de masas sobre los enlaces químicos presentes en las moléculas. Debido a que las moléculas orgánicas tienden a ser muy grande, el número de estructuras posibles para estas moléculas tiende a ser enorme. DENDRAL aborda el problema de este gran espacio de búsqueda mediante la aplicación del conocimiento heurístico de expertos químicos al problema elucidación de la estructura. métodos de DENDRAL resultaron notablemente eficaz, la búsqueda de forma rutinaria la estructura correcta de millones de posibilidades después de sólo unos pocos ensayos. El enfoque ha tenido tanto éxito que los descendientes extensiones de DENDRAL se utilizan actualmente en los laboratorios químicos y farmacéuticos en todo el mundo.

Mientras que DENDRAL fue uno de los primeros programas para utilizar eficazmente los conocimientos de dominio específico para lograr un rendimiento de expertos, MYCIN estableció la metodología de sistemas expertos contemporáneos (Buchanan y Shortliffe 1984). MYCIN utiliza el conocimiento de expertos médicos para diagnosticar y prescribir tratamiento para la meningitis espinal y las infecciones bacterianas de la sangre. MYCIN, desarrollado en Stanford a mediados de la década de 1970, fue uno de los primeros programas para abordar los problemas de razonamiento con información incierta o incompleta.

MYCIN proporciona explicaciones claras y lógicas de su razonamiento, utiliza una estructura de control adecuadas para el dominio del problema específico, y los criterios identificados para evaluar de forma fiable su rendimiento. Muchas de las técnicas de desarrollo de sistemas expertos actualmente en uso se desarrollaron por primera vez en el proyecto MYCIN (capítulo 8).

Otros sistemas expertos tempranos incluyen el programa PROSPECTOR para determinar la probable ubicación y tipo de depósitos de mineral basado en información geológica acerca de un sitio (Duda et al., 1979a, 1979b), el programa INTERNO para realizar el diagnóstico en el área de la medicina interna, la Medición de Echados asesor para la interpretación de los resultados de los registros de perforación de pozos de petróleo (Smith y Baker, 1983), y XCON para la configuración de los ordenadores VAX. XCON fue desarrollado en 1981, y en un tiempo cada VAX vendidos por Digital EquipmentCorporation fue configurado por el software. Otros numerosos sistemas expertos actualmente están resolviendo problemas en áreas como la medicina, la educación, los negocios, el diseño y la ciencia (Waterman 1986, Durkin, 1994). Ver también los procedimientos actuales de las aplicaciones Inovative de la Inteligencia Artificial (IAAI) Conferencias.

Es interesante observar que la mayoría de los sistemas expertos se han escrito para dominios de nivel relativamente especializados, expertos. Estos dominios son generalmente bien estudiados y claramente han definido las estrategias de resolución de problemas. Los problemas que dependen de una noción más vagamente definido de "sentido común" son mucho más difíciles de resolver por estos medios. A pesar de la promesa de los sistemas expertos, sería un error de sobreestimar la capacidad de esta tecnología. Deficiencias actuales incluyen:

1. Dificultad para capturar el conocimiento "profundo" del dominio del problema. MYCIN, por ejemplo, carece de cualquier conocimiento real de la fisiología humana. No sabe lo que hace la sangre o la función de la médula espinal. Cuenta la leyenda que una vez, cuando la selección de un medicamento para el tratamiento de la meningitis, MYCIN se le preguntó si el paciente estaba embarazado, a pesar de que le habían dicho que el paciente era un hombre. Si esto realmente ocurrió o no, sirve para ilustrar la estrechez potencial del conocimiento en sistemas expertos.

2. La falta de robustez y flexibilidad. Si los seres humanos son presentados con una instancia de un problema que no pueden resolver de inmediato, por lo general pueden regresar a un examen de los primeros principios y llegar a alguna estrategia para atacar el problema. Los sistemas expertos generalmente carecen de esta habilidad.

3. Incapacidad para proporcionar explicaciones profundas. Debido a que los sistemas expertos carecen de un conocimiento profundo de sus dominios del problema, sus explicaciones se limitan generalmente a una descripción de los pasos que tomaron en la búsqueda de una solución. Por ejemplo, a menudo no pueden decir "por qué" un cierto enfoque fue adoptado.

4. Las dificultades en la verificación. A pesar de la corrección de cualquier sistema informático grande es difícil de probar, los sistemas expertos son particularmente difíciles de verificar. Este es un problema grave, ya que la tecnología de los sistemas expertos se aplica a las aplicaciones críticas tales como el control del tráfico aéreo, las operaciones del reactor nuclear, y sistemas de armas.

5. pequeño que aprende de la experiencia. Sistemas expertos actuales son hechos a mano; una vez que el sistema se completa, su rendimiento no mejorará sin más la atención de sus programadores, lo que lleva a las dudas sobre la inteligencia de este tipo de sistemas.

A pesar de estas limitaciones, los sistemas expertos han demostrado su valor en un número de aplicaciones importantes. Los sistemas expertos son un tema importante en este texto y se discuten en los capítulos 7 y 8. Las aplicaciones actuales a menudo se pueden encontrar en las actas de las aplicaciones Inovative de la Inteligencia Artificial (IAAI) conferencias.

**1.2.4Comprensión de Lenguaje Natural y Semántica**

Uno de los objetivos de larga data de la inteligencia artificial es la creación de programas que son capaces de entender y generar lenguaje humano. No sólo la capacidad de utilizar y comprender el lenguaje natural parece ser un aspecto fundamental de la inteligencia humana, sino también su automatización exitosa tendría un impacto increíble en la facilidad de uso y la eficacia de los propios ordenadores. Mucho esfuerzo se ha puesto en escribir programas que comprenden el lenguaje natural. Aunque estos programas han logrado el éxito en contextos restringidos, sistemas que pueden utilizar lenguaje natural con la flexibilidad y generalidad que caracterizan el habla humana están más allá de las metodologías actuales.

La comprensión del lenguaje natural implica mucho más que analizar frases en sus partes individuales del habla y mirando esas palabras en un diccionario. comprensión real depende en el fondo un amplio conocimiento sobre el dominio del discurso y los idiomas utilizados en ese dominio, así como la capacidad de aplicar el conocimiento contextual en general para resolver las omisiones y ambigüedades que son una parte normal de la voz humana.

Consideremos, por ejemplo, las dificultades para mantener una conversación sobre el béisbol con una persona que entiende Inglés, pero no sabe nada acerca de las reglas, jugadores, o la historia del juego. ¿Podría esta persona posiblemente entender el significado de la frase:

"Con ninguno en la parte superior de la novena y la carrera de la ventaja en la segunda, el gerente llamó a su alivio del corral de toros"? A pesar de que todas las palabras en la frasepueden entenderse de forma individual, esta frase sería un galimatías para igualar el ventilador más inteligente del no-béisbol.

La tarea de recolección y organización de este conocimiento de fondo de tal manera que se puede aplicar a la comprensión del lenguaje constituye el principal problema en la automatización de la comprensión del lenguaje natural. En respuesta a esta necesidad, los investigadores han desarrollado muchas de las técnicas para la estructuración de significado semántico utilizado a través de la inteligencia artificial (capítulos 7 y 15).

Debido a las enormes cantidades de conocimientos necesarios para la comprensión del lenguaje natural, la mayoría del trabajo se realiza en las áreas problemáticas, especializadas bien entendidos. Uno de los primeros programas de aprovechar esta metodología de "micro mundo" era SHRDLU de Winograd, un sistema de lenguaje natural que podría "conversar", sobre una sencilla configuración de los bloques de diferentes formas y colores (Winograd, 1973). SHRDLU podía responder a las preguntas tales como "¿de qué color es el bloque en el cubo azul?", Así como las acciones del plan, tales como el "movimiento de la pirámide roja en el ladrillo verde". Los problemas de este tipo, que implican la descripción y manipulación de arreglos sencillos de bloques, aparecieron con una frecuencia sorprendente en las primeras investigaciones de AI y se conocen como "bloques de problemas mundiales".

A pesar del éxito de SHRDLU en conversar sobre los arreglos de bloques, sus métodos no se generalizan a partir de ese dominio. Las técnicas de representación utilizados en el programa eran demasiado simples para capturar la organización semántica de los dominios más ricas y complejas de una manera útil. Gran parte del trabajo actual en la comprensión del lenguaje natural se dedica a la búsqueda de los formalismos de representación que son suficientes para ser utilizado en una amplia gama de aplicaciones sin embargo, adaptarse bien a la estructura específica de un determinado dominio general. Una serie de técnicas diferentes (muchos de los cuales son extensiones o modificaciones de las redes semánticas) son explorados para este propósito y se utiliza en el desarrollo de programas que pueden comprender el lenguaje natural en dominios de conocimiento limitados, pero interesantes. Por último, en la investigación actual (Marcus 1980, Manning y Schütze, 1999 Jurafsky y Martin 2009) modelos estocásticos, que describen cómo "se producen" las palabras y las estructuras de la lengua en uso, se emplean para caracterizar tanto la sintaxis y la semántica. Comprensión computacional completa del lenguaje, sin embargo, permanece más allá del estado actual de la técnica.

**1.2.5Modelado de Rendimiento Humano**

Aunque gran parte de la discusión anterior utiliza la inteligencia humana como punto de referencia en la consideración de la inteligencia artificial, no se sigue que los programas deben patrón de sí mismos después de la organización de la mente humana. De hecho, muchos programas de IA están diseñados para resolver algún problema útil sin tener en cuenta sus similitudes con la arquitectura mental humana. Incluso los sistemas expertos, mientras que deriva gran parte de su conocimiento de los expertos humanos, no realmente tratar de simular procesos internos humanos con problemas mentales para resolver problemas.

Si el rendimiento es el único criterio por el que será juzgado un sistema, puede haber pocas razones para tratar de simular los métodos de resolución de problemas humanos; de hecho, los programas que se llevan a enfoques no humanos a la solución de problemas (ajedrez) son a menudo más éxito que sus homólogos humanos. Aun así, el diseño de sistemas que modelan explícitamente los aspectos de la actuación humana es una zona fértil de la investigación, tanto en la IA y la psicología.

El modelado del rendimiento humano, además de proporcionar AI con gran parte de su metodología básica, ha demostrado ser una poderosa herramienta para formular y probar las teorías de la cognición humana. Las metodologías de resolución de problemas desarrollados por los informáticos psicólogos han dado una nueva metáfora para explorar la mente humana.

En lugar de fundición teorías de la cognición en el lenguaje impreciso utilizado en la investigación temprana o abandonar el problema de describir el funcionamiento interno de la mente humana en su totalidad (según lo sugerido por los conductistas), muchos psicólogos han adoptado el lenguaje y la teoría de la informática para formular modelos de la inteligencia humana. No sólo estas técnicas proporcionan un nuevo vocabulario para describir la inteligencia humana, sino también las implementaciones informáticas de estas teorías ofrecen los psicólogos la oportunidad de someter a prueba en, crítica y refinar sus ideas (Luger 1994; véase el Diario y las conferencias de la Sociedad de Ciencia Cognitiva).

**1.2.6Planificación y Robótica**

La investigación en la planificación comenzó como un esfuerzo para diseñar robots que pudieran llevar a cabo sus tareas con cierto grado de flexibilidad y capacidad de respuesta al mundo exterior. En pocas palabras, la planificación asume un robot que es capaz de realizar ciertas acciones atómicas. Se trata de encontrar una secuencia de aquellas acciones que lograrán alguna tarea de nivel superior, como se mueve a través de una habitación llena de obstáculos. La planificación es un problema difícil para una serie de razones, no menos importante de los cuales es el tamaño del espacio de posibles secuencias de movimientos. Incluso un robot extremadamente simple es capaz de generar un gran número de posibles secuencias de movimiento. Imaginemos, por ejemplo, un robot que puede moverse hacia adelante, atrás, derecha o izquierda, y cuenta el número de formas que posiblemente robot puede moverse por una habitación diferente. Supóngase también que hay obstáculos en la habitación y que el robot debe seleccionar una ruta que se mueve alrededor de ellos de alguna manera eficiente.

Escribir un programa que puede descubrir la mejor ruta en estas circunstancias, sin ser abrumados por el gran número de posibilidades, requiere técnicas sofisticadas para representar el conocimiento espacial y el control de búsqueda a través de los entornos posibles.

Uno de los métodos que los seres humanos utilizan en la planificación es la descomposición jerárquica problema.

Si usted está planeando un viaje desde Albuquerque a Londres, por lo general, trata los problemas de la organización de un vuelo, llegar al aeropuerto, haciendo conexiones aéreas, y la búsqueda de transporte terrestre en Londres por separado, a pesar de que todos ellos son parte de un plan general más grande. Cada uno de éstos se puede descomponer en subproblemas más pequeños, tales como la búsqueda de un mapa de la ciudad, la negociación del sistema de metro, y la búsqueda de un bar decente. No sólo este enfoque limita efectivamente el tamaño del espacio que debe ser buscado, pero también es compatible con el ahorro de subplanes se utilizan con frecuencia para uso futuro.

Mientras que los humanos planean sin esfuerzo, la creación de un programa de ordenador que puede hacer lo mismo es un reto difícil. Una tarea aparentemente simple, como romper un problema en subproblemas independientes requiere realmente sofisticados heurística y un amplio conocimiento sobre el dominio de planificación. Determinar qué planes secundarios deben ser guardados y cómo pueden ser generalizados para su uso futuro es un problema igualmente difícil. Un robot que realiza a ciegas una secuencia de acciones sin responder a los cambios en su entorno o ser capaz de detectar y corregir errores en su propio plan difícilmente podrían ser considerados inteligente. Un robot no puede tener sensores adecuados para localizar a todos los obstáculos en el camino de una trayectoria proyectada.

Un robot de este tipo debe comenzar a moverse a través de la habitación sobre la base de lo que se ha "percibido" y corregir su trayectoria que se detectan otros obstáculos. La organización de los planes de una manera que permite una respuesta a las cambiantes condiciones ambientales es un problema importante para la planificación (Lewis y Luger de 2000, Thrun et al., 2007).

Por último, la robótica fue una de las áreas de investigación en IA que produjeron muchas de las ideas que apoyan la resolución de problemas orientado a agentes (Sección 1.1.4). Frustrado por tanto la complejidad de mantener el gran espacio de representación, así como el diseño de algoritmos de búsqueda para la planificación tradicional, los investigadores, incluyendo Agre y Chapman (1987), Brooks (1991a), Thrun et al. (2007), actualizado el problema en términos de la interacción de múltiples agentes semiautónomos. Cada agente es responsable de su propia parte de la tarea a través de su problema y la coordinación surgiría la solución más grande.

Planificación de la investigación ahora se extiende mucho más allá de los dominios de la robótica, para incluir la coordinación de cualquier complejo conjunto de tareas y objetivos. Planificadores modernos se aplican a los agentes (Nilsson 1994), así como para el control de aceleradores de haces de partículas (Klein et al. 1999, 2000).

**1.2.7Lenguajes y entornos de AI**

Algunos de los productos secundarios más importantes de la investigación de la inteligencia artificial han sido los avances en los lenguajes de programación y entornos de desarrollo de software. Por una serie de razones, incluyendo el tamaño de muchos programas de aplicación de IA, la importancia de una metodología de creación de prototipos, la tendencia de los algoritmos de búsqueda para generar espacios enormes, y la dificultad de predecir el comportamiento de los programas heurísticamente impulsadas, los programadores de AI han visto obligados a desarrollar un potente conjunto de metodologías de programación.

Entornos de programación incluyen técnicas de conocimiento-estructurantes como la programación orientada a objetos. Los lenguajes de alto nivel, tales como Lisp y Prolog, que apoyan el desarrollo modular, ayudan a controlar el tamaño y la complejidad del programa. Paquetes de trazas permiten a un programador para reconstruir la ejecución de un algoritmo complejo y hacen posible desentrañar las complejidades de búsqueda heurística. Sin este tipo de herramientas y técnicas, es dudoso que muchos sistemas importantes de IA podrían haber sido construidos.

Muchas de estas técnicas son ahora herramientas estándar para la ingeniería de software y tienen poca relación con el núcleo de la teoría de la IA. Otros, como la programación orientada a objetos, son de interés teórico y práctico significativo. Por último, muchos algoritmos de IA también se construyen ahora en lenguajes informáticos más tradicionales, tales como C ++ y Java.

Los lenguajes de programación desarrollados para la inteligencia artificial están íntimamente ligados a la estructura teórica del campo. Hemos construido muchas de las estructuras de representación presneted en este libro en Prolog, Lisp y Java, que estarán disponibles en la Luger y Stubblefield (2009) y en Internet. En este libro nos mantenemos al margen de los debates religiosos sobre sus méritos relativos de diferentes idiomas. Más bien, se adhieren al adagio de "el buen profesional sabe todas sus herramientas."

**1.2.8Aprendizaje automático**

El aprendizaje ha seguido siendo un ámbito problemático para la IA. La importancia del aprendizaje, sin embargo, está fuera de discusión, especialmente en lo que esta capacidad es uno de los componentes más importantes de la conducta inteligente. Un sistema experto puede realizar cálculos extensos y costosos para resolver un problema. A diferencia de un ser humano, sin embargo, si se le da el mismo o un problema similar por segunda vez, por lo general no se acuerda de la solución. Se lleva a cabo la misma secuencia de cálculos de nuevo. Esto es cierto el segundo, tercero, cuarto, y cada vez que resuelve ese problema, apenas el comportamiento de un solucionador de problemas inteligente. La solución obvia a este problema es que los programas para aprender por su cuenta, ya sea por experiencia, la analogía, ejemplos, siendo "dijo" qué hacer, o premiados o castigados en función de los resultados.

A pesar de que el aprendizaje es un área difícil, hay varios programas que apuntan a que no es imposible. Un programa para la primera es AM, el matemático automatizado, diseñado para descubrir las leyes matemáticas (Lenat 1977, 1982). Inicialmente dados los conceptos y axiomas de la teoría de conjuntos, AM fue capaz de inducir dichos conceptos matemáticos importantes como cardinalidad, la aritmética de enteros, y muchos de los resultados de la teoría de números. AM conjeturó nuevos teoremas mediante la modificación de su base de conocimientos actual y utiliza la heurística para perseguir el "mejor" de una serie de posibles alternativas teoremas. Más recientemente, Cotton et al. (2000) diseñaron un programa que inventa automáticamente secuencias de números enteros "interesantes".

A principios influyente obra también incluye la investigación de Winston en la inducción de conceptos estructurales tales como "arco" de un conjunto de ejemplos en el mundo bloques (Winston 1975a). El algoritmo ID3 ha demostrado tener éxito en el aprendizaje de patrones generales de ejemplos (Quinlan 1986a). Meta-DENDRAL aprende las reglas para la interpretación de los datos de masa espectrográfico en química orgánica a partir de ejemplos de datos sobre los compuestos de estructura conocida.

Tiresias, un ser inteligente "front-end" de sistemas expertos, convierte asesoramiento de alto nivel en nuevas reglas para su base de conocimientos (Davis 1982). Hacker maquina planes para llevar a cabo manipulaciones bloques mundiales a través de un proceso iterativo de la elaboración de un plan, probándola, y corregir los defectos descubiertos en el plan de candidato (Sussman, 1975). El trabajo en el aprendizaje basado en la explicación ha demostrado la eficacia de los conocimientos previos en el aprendizaje (Mitchell et al., 1986, DeJong y Mooney 1986). También hay ahora muchos importantes modelos biológicos y sociológicos del aprendizaje; se revisan estos en los de aprendizaje y de aprendizaje emergente capítulos conexionistas.

El éxito de los programas de aprendizaje automático sugiere la existencia de un conjunto de principios generales de aprendizaje que permitirán la construcción de programas con la capacidad de aprender en los dominios realistas. Presentamos varios enfoques de aprendizaje en la Sección IV.

**1.2.9Representaciones alternativas: Redes Neuronales y Algoritmos Genéticos**

La mayor parte de las técnicas que se presentan en este libro uso AI representados de manera explícita el conocimiento y cuidadosamente diseñado algoritmos de búsqueda para poner en práctica la inteligencia. Un enfoque muy diferente busca construir programas inteligentes que utilizan modelos que son paralelos a la estructura de las neuronas en el cerebro humano o los patrones cambiantes que se encuentran en los algoritmos genéticos y la vida artificial.

Un esquema simple de una neurona (Figura 1.2) consta de un cuerpo celular que tiene un número de salientes ramificados, llamados dendritas, y una sola rama llamada el axón. Las dendritas reciben señales de otras neuronas. Cuando estos impulsos combinados superan un cierto umbral, los fuegos de neuronas y un impulso, o pico, pasa por el axón. Ramas en el extremo de la forma axón sinapsis con las dendritas de otras neuronas. La sinapsis es el punto de contacto entre las neuronas; sinapsis pueden ser excitadoras o inhibidoras, ya sea añadiendo a la suma de las señales que llegan a la neurona o restando de ese total.

Esta descripción de una neurona es excesivamente simple, pero captura aquellas características que son relevantes para modelos neuronales de cálculo. En particular, cada unidad de cálculo calcula una función de sus entradas y pasa el resultado a lo largo de las unidades conectadas en la red:los resultados finales son producidos por el procesamiento en paralelo y distribuido de esta red de conexiones neuronales y pesos de umbral. Arquitecturas neuronales son atractivos mecanismos para la aplicación de la inteligencia para un número de razones. los programas de IA tradicionales pueden ser frágiles y muy sensibles al ruido.

La inteligencia humana es mucho más flexible y bueno en la interpretación de entrada con ruido, como por ejemplo una cara en una habitación oscura o una conversación en una fiesta ruidosa. Arquitecturas neuronales, porque capturan conocimiento en un gran número de unidades de grano fino distribuidos sobre una red, parecen tener más potencial para hacer coincidir parcialmente datos ruidosos e incompletos.

Con los algoritmos genéticos y la vida artificial nos desarrollamos nuevas soluciones a los problemas de los componentes de las soluciones anteriores. Los operadores genéticos, como el cruce y mutación, al igual que sus equivalentes genéticos en el mundo natural, el trabajo para producir, para cada nueva generación, cada vez mejor el potencial de las soluciones de problemas. La vida artificial produce su nueva generación como una función de la "calidad" de sus vecinos en las generaciones anteriores.

Ambas arquitecturas neuronales y algoritmos genéticos proporcionan un modelo natural para el paralelismo, porque cada neurona o segmento de una solución es una unidad independiente.

Hillis (1985) ha comentado sobre el hecho de que los seres humanos se vuelven más rápidos en una tarea a medida que adquieren más conocimiento, mientras que las computadoras tienden a reducir la velocidad. Esta desaceleración se debe al coste de forma secuencial buscar una base de conocimientos; una arquitectura masivamente paralela al igual que el cerebro humano no sufriría de este problema. Por último, algo es intrínsecamente atractivo de acercarse a los problemas de la inteligencia desde un punto de vista neural o genética. Después de todo, el cerebro evolucionado logra inteligencia y lo hace utilizando una arquitectura neural. Presentamos las redes neuronales, algoritmos genéticos, y la vida artificial, en los capítulos 10 y 11.

