# INTELIGENCIA ARTIFICIAL

# Unidad 1: Inteligencia artificial y agentes inteligentes

## Inteligencia Artificial

### Que es la IA

Las definiciones que aparecen en la parte superior se refieren a procesos mentales y al razonamiento, mientras que las de la parte inferior aluden a la conducta. Las definiciones de la izquierda miden el éxito en términos de la fidelidad en la forma de actuar de los humanos, mientras que las de la derecha toman como referencia un concepto ideal de inteligencia, que llamaremos racionalidad. Un sistema es racional si hace «lo correcto», en función de su conocimiento. El enfoque centrado en el comportamiento humano debe ser una ciencia empírica, que incluya hipótesis y confirmaciones mediante experimentos. El enfoque racional implica una combinación de matemáticas e ingeniería.

Texto, Tabla

Descripción generada automáticamente con confianza media

#### Comportamiento Humano: la prueba de Turing

La **Prueba de Turing**, propuesta por Alan Turing (1950), se diseñó para proporcionar una definición operacional y satisfactoria de inteligencia. En vez de proporcionar una lista de cualidades necesarias para obtener inteligencia artificialmente, él sugirió una prueba basada en la incapacidad de diferenciar entre entidades inteligentes indiscutibles y seres humanos. El computador supera la prueba si un evaluador humano no es capaz de distinguir si las respuestas, a una serie de preguntas planteadas, son de una persona o no. Programar un computador para que supere la prueba requiere un trabajo considerable porque el computador debería poseer las siguientes capacidades:

1. **Procesamiento de lenguaje natural** que le permita comunicarse satisfactoriamente en inglés.
2. **Representación del conocimiento** para almacenar lo que se conoce o siente.
3. **Razonamiento automático** para utilizar la información almacenada para responder a preguntas y extraer nuevas conclusiones.
4. **Aprendizaje automático** para adaptarse a nuevas circunstancias y para detectar y extrapolar patrones.

La Prueba de Turing evitó deliberadamente la interacción físicadirecta entre el evaluador y el computador, dado que para medir la inteligencia es innecesario simular físicamente a una persona. La Prueba Global de Turing incluye una señal de vídeo que permite al evaluador valorar la capacidad de percepción del evaluado, y también le da la oportunidad al evaluador de pasar objetos físicos. Para superar la Prueba Global de Turing el computador debe estar dotado de

1. **Visión computacional** para percibir objetos.
2. **Robótica** para manipular y mover objetos.

Estas seis disciplinas abarcan la mayor parte de la IA.

#### Pensar como Humano: el enfoque cognitivo

Para poder decir que un programa dado piensa como un humano, es necesario contar con un mecanismo para determinar cómo piensan los humanos. Se puede hacer mediante introspección (intentando atrapar nuestros propios pensamientos conforme éstos van apareciendo) y mediante experimentos psicológicos. Una vez se cuente con una teoría lo suficientemente precisa sobre cómo trabaja la mente, se podrá expresar esa teoría en la forma de un programa de computador. Si los datos de entrada/salida del programa y los tiempos de reacción son similares a los de un humano, existe la evidencia de que algunos de los mecanismos del programa se pueden comparar con los que utilizan los seres humanos. En el campo interdisciplinario de la **ciencia cognitiva** convergen modelos computacionales de IA y técnicas experimentales de psicología intentando elaborar teorías precisas y verificables sobre el funcionamiento de la mente humana.

#### Pensamiento Racional: el enfoque de las Leyes del Pensamiento

Aristóteles intentó codificar la «manera correcta de pensar», es decir, un proceso de razonamiento irrefutable. Sus **silogismos** son esquemas de estructuras de argumentación mediante las que siempre se llega a conclusiones correctas si se parte de premisas correctas. Estas leyes de pensamiento supuestamente gobiernan la manera de operar de la mente; su estudio fue el inicio de un campo llamado **lógica**.

Estudiosos de la lógica desarrollaron una notación precisa para definir sentencias sobre todo tipo de elementos del mundo y especificar relaciones entre ellos. Ya en 1965 existían programas que, en principio, resolvían *cualquier* problema resoluble descrito en notación lógica. La llamada tradición **logista** dentro del campo de la inteligencia artificial trata de construir sistemas inteligentes a partir de estos programas.

#### Actuar de Forma Racional: el enfoque del agente racional

Un **agente** es algo que razona. Pero de los agentes informáticos se espera que tengan otros atributos que los distingan de los programas convencionales (controles autónomos, percepción del entorno, persistencia en el tiempo, adaptación a cambios, que sean capaces de alcanzar objetivos diferentes).

Un **agente racional** es aquel que actúa con la intención de alcanzar el mejor resultado o, cuando hay incertidumbre, el mejor resultado esperado.

En el caso del enfoque de la IA según las «leyes del pensamiento», todo el énfasis se pone en hacer inferencias correctas. La obtención de estas inferencias correctas puede formar *parte* de lo que se considera un agente racional, ya que una manera racional de actuar es llegar a la conclusión lógica de que, si una acción dada permite alcanzar un objetivo, hay que llevar a cabo dicha acción. Sin embargo, el efectuar una inferencia correcta no depende siempre de la *racionalidad*, ya que existen situaciones para las que no hay nada correcto que hacer y en las que hay que tomar una decisión. Existen también formas de actuar racionalmente que no implican realizar inferencias.

Todas la habilidades que se necesitan en la Prueba de Turing deben permitir emprender acciones racionales. Por lo tanto, es necesario contar con la capacidad para representar el conocimiento y razonar basándonos en él, porque ello permitirá alcanzar decisiones correctas en una amplia gama de situaciones.

### ¿Por qué es interesante la IA?

#### Sentido de la Identidad

El hombre se ha aplicado a sí mismo el nombre científico de ***homo sapiens***(hombre sabio o "capaz de conocer") como una valoración de la trascendencia de nuestras habilidades mentales tanto para la vida cotidiana como en nuestro propio sentido de la identidad.

#### Comprensión de Entidades Inteligentes

Los esfuerzos en el campo de la IA se enfocan en lograr la compresión de entidades inteligentes. Por eso, una de las razones de su estudio es aprender más sobre nosotros mismos. A diferencia de la filosofía y de la psicología, que también se ocupan de la inteligencia, los esfuerzos de la IA están encaminados tanto a la comprensión de entidades inteligentes como su construcción. Otra razón por la que se estudia la IA es que las entidades inteligentes así construidas son interesantes y útiles por derecho propio.

#### Es un problema Complejo pero

El problema que aborda la IA es uno de los más complejos, pero a diferencia de la investigación en torno al desplazamiento a mayor velocidad que la luz o de un dispositivo antigravitatorio, el investigador del campo de la IA cuenta con pruebas contundentes de que tal búsqueda es factible. Todo lo que este investigador tiene que hacer es mirarse en un espejo para tener ante sí un ejemplo de sistema inteligente.

#### El estudio de la inteligencia es una de las disciplinas más antiguas

Por más de 2000 años los filósofos se han preguntado cómo se ve, aprende, recuerda y razona. La llegada de las computadoras a principio de la década de los cincuenta permitió pasar de la especulación, a nivel de charlas de café, en torno a estas facultades mentales a su abordaje mediante una auténtica disciplina teórica y experimental.

### Fundamentos (De otras ciencias)

Si bien la IA es un campo joven, es heredera de diversas ideas, puntos de vista y técnicas de otras disciplinas.

|  |  |
| --- | --- |
| **Disciplina** | **Aporte** |
| **Filosofía** | * **Modelos de razonamiento y lógica:** Los filósofos, como Aristóteles, sentaron las bases para la lógica formal y el razonamiento deductivo, principios que son fundamentales para el diseño de sistemas de razonamiento en la IA, como sistemas expertos y sistemas de inferencia. * **Dualismo y materialismo:** Influyeron en las concepciones de cómo la mente y la inteligencia podrían ser representadas y replicadas en máquinas.   Dada una mente física que gestiona conocimiento, el siguiente problema es establecer  las fuentes de este conocimiento.   * **Movimiento empírico:** proporcionó enfoques y conceptos clave para abordar la adquisición de conocimiento, el razonamiento y la toma de decisiones en los sistemas de IA. * **Principio de inducción:** influyó en el desarrollo de sistemas de aprendizaje automático y la toma de decisiones basada en datos. La inducción es un proceso mediante el cual se infieren patrones generales o reglas a partir de ejemplos específicos. * **Positivismo Lógico:** esta doctrina sostiene que todo el conocimiento se puede caracterizar mediante teorías lógicas relacionadas, en última instancia, con **sentencias de observación** que corresponden a estímulos sensoriales. * **Teoría de la Confirmación:** intenta explicar cómo el conocimiento se * obtiene a partir de la experiencia. |
| **Matemáticas** | Para pasar a una ciencia formal se necesitaba una formulación matemática en:   1. Lógica:  * Lógica proposicional o Booleana * Lógica de Primer Orden: extiende la lógica de Boole para incluir objetos y relaciones. Se utiliza hoy como el sistema más básico de representación de conocimiento. * Teoría de Referencia: enseña cómo relacionar objetos de una lógica con objetos del * mundo real.  1. Computación:  * Avances en Algoritmos. * Teorema de Incompletitud: en cualquier lenguaje con capacidad suficiente para expresar las propiedades de los números naturales, existen aseveraciones verdaderas de las cuales no es posible decidir su validez mediante ningún algoritmo. * Intratabilidad: un problema es intratable si el tiempo necesario para la resolución de casos particulares de dicho problema crece exponencialmente con el tamaño de dichos casos. * NP-Completitud: es un método para reconocer problemas intratables.  1. Probabilidad |
| **Economía** | * Teoría de la decisión * Teoría de Juegos * Investigación Operativa * Satisfacción. |
| **Neurociencia** | Es el estudio del sistema neurológico, y en especial del cerebro. En la actualidad se dispone de información sobre la relación existente entre las áreas del cerebro y las partes del cuerpo humano que controlan o de las que reciben impulsos sensoriales. |
| **Psicología** | * **Conductismo:** estudio exclusivo de mediciones objetivas de percepciones (o *estímulos*) sobre animales y de las acciones resultantes (o *respuestas*). Construcciones mentales como conocimientos, creencias, objetivos y pasos en un razonamiento quedaron descartadas por ser consideradas «psicología popular» no científica. * **Psicología cognitiva:** su característica es la conceptualización del cerebro como un dispositivo de procesamiento de información. * **Ciencia Cognitiva:** surge del desarrollo del modelo computacional. Se presentaron artículos que mostraban cómo se podían utilizar los modelos informáticos para modelar la psicología de la memoria, el lenguaje y el pensamiento lógico, respectivamente. |
| **Computación** | Para que la inteligencia artificial pueda llegar a ser una realidad se necesitan inteligencia y un artefacto: el computador. Desde mediados del siglo pasado, cada generación de dispositivos *hardware* ha conllevado un aumento en la velocidad de proceso y en la capacidad de almacenamiento, asi como una reducción de precios. |
| **Teoría de Control** | Se visualiza el comportamiento determinista como algo emergente de un mecanismo regulador que intenta minimizar el error (la diferencia entre el estado presente y el estado objetivo). La teoría de control moderna tiene como objetivo el diseño de sistemas que maximizan una **función objetivo** en el tiempo. Lo cual se asemeja ligeramente a nuestra visión de lo que es la IA: diseño de sistemas que se comportan de forma óptima. |
| **Lingüística** | La lingüística moderna y la IA «nacieron», al mismo tiempo y maduraron juntas, solapándose  en un campo híbrido llamado **lingüística computacional** o **procesamiento del lenguaje natural**. El entendimiento del lenguaje requiere la comprensión de la materia bajo estudio y de su contexto, y no solamente el entendimiento de la estructura de las sentencias. |

### Breve historia y algunos hitos

|  |  |
| --- | --- |
| **Año** | **Avance** |
| **1936** | Alan Turing crea la “Máquina de Turing”, la base de las computadoras. |
| **1943** | Warren McCulloch y Walter Pitts son considerados pioneros en la IA, desarrollando un modelo de neuronas artificiales basado en conocimientos de fisiología cerebral, lógica proposicional y teoría de la computación de Turing. Crearon un modelo de neuronas que podían estar "activadas" o "desactivadas", con activación estimulada por otras neuronas. Vieron el estado de una neurona como equivalente a una proposición con estímulos adecuados. Demostraron que las funciones de cómputo y los conectores lógicos podían implementarse con redes de neuronas. |
| **1949** | Donald Hebb propuso la regla de aprendizaje Hebbiano para ajustar conexiones neuronales según la estimulación, idea que sigue vigente. |
| **1950** | Presentó la prueba de Turing (una prueba para determinar si una máquina es inteligente como un humano), el aprendizaje automático, los algoritmos genéricos y el aprendizaje por refuerzo. |
| **1951** | Marvin Minsky y Dean Edmonds construyeron el primer computador basado en red neuronal en 1951, llamado SNARC. |
| **1952** | McCarthy organiza un taller en Darmouth a la que asistieron científicos que trabajaban en temas relacionados con "máquinas pensantes". Allen Newell y Herbert Simon captaron la atención con su programa de razonamiento llamado Teórico Lógico (TL), que demostró teoremas matemáticos. Acá se acuñó el nombre de “Inteligencia Artificial”. |
| **1956-1969** | Optimismo exagerado. Se resolvieron problemas como la demostración de teoremas y se logró que las máquinas superaran a los humanos en algunos juegos simples. |
| **1966-1973** | Se intentaron atacar problemas del dominio del sentido común o aquellos que necesitan de conocimiento general, como los relacionados con el lenguaje, los esfuerzos no fueron suficientes. Esta situación llevó a un desanimo generalizado y a la pérdida de inversiones. |
| **1969-1979** | Se centró en el desarrollo de métodos de búsqueda generales y en el uso de conocimiento específico del dominio para resolver problemas más complejos. |
| **1980 – Presente** | La IA resurgió de la mano de los Sistemas Expertos. Desde ese momento el área no deja de crecer. Los sistemas de visión artificial y los de reconocimiento de patrones han provisto innumerables soluciones en los últmos años. La cantidad de datos disponibles, los recursos computacionales y los avances como el de las redes neuronales profundas (*deep learning*) han formado un círculo virtuoso donde tanto las empresas como la comunidad científica invierten cada vez más recursos. |

### Corrientes

Desde los inicios de la IA se plantearon dos enfoques diferentes, uno por parte de los investigadores del MIT y otro por los del Carnegie-Mellon. Estos caminos se fueron alejando hasta llegar a plantearse un distanciamiento que caracterizó las investigaciones en IA de los veinte años siguientes y que fue sin lugar a dudas muy perjudicial por el atraso que provocó en algunas líneas de investigación.

#### Corriente conexionista - subsimbólica - ascendente.

La escuela liderada por los investigadores de la Universidad de Carnegie-Mellon proponía desarrollar modelos del comportamiento humano a partir de elementos que simularan en todo lo posible al cerebro con un enfoque claramente cognitivo. Así iniciaron una corriente denominada "de aproximaciones subsimbólicas" que sigue un diseño de modelo "ascendente", comenzando en el nivel más bajo y operando hacia niveles superiores. Según esta corriente, para concebir máquinas inteligentes hay que seguir muchos de los pasos evolutivos que acompañaron el desarrollo de la inteligencia en los seres vivos. De esta forma hay que comenzar por replicar la capacidad de procesamiento de señales y subir por la cadena evolutiva en pasos sucesivos, construyendo progresivos niveles superiores de inteligencia que constituirán las bases sólidas para los pasos siguientes. Al amparo de esta propuestase desarrollaron los modelos conexionistas, redes neuronales inspiradas en los modelos biológicos.

#### Corriente tradicional - simbólica - descendente.

La segunda escuela, liderada por Minsky y McCarthy del MIT orientó su actividad a obtener comportamientos "inteligentes" sin procurar que la estructura de los componentes tuviese semejanza con sus equivalentes biológicos y llegó a su mayor auge en los primeros 20 años de la IA. Esta escuela iniciada en el MIT fue denominada "tradicional", "de procesamiento de símbolos" o "de diseño descendente" y se apoyó en representar el conocimiento mediante sentencias declarativas y en la deducción de sus consecuencias a partir de la aplicación de reglas de inferencia. Sin embargo, poco a poco se fue comprobando que muchas de las soluciones propuestas eran apropiadas para modelos pequeños, pero completamente ineficientes o inaplicables cuando se pretendía resolver problemas de tamaño real. Los investigadores se encontraron con que sus sistemas sucumbían ante la creciente longitud y complejidad de su programación.

* 1. **Agentes Inteligentes**

Agentes y su entorno

Un agente es cualquier cosa capaz de percibir su medioambiente con la ayuda de sensores y actuar en ese medio utilizando actuadores. Se trabajará con la hipótesis general de que cada agente puede percibir sus propias acciones (pero no siempre sus efectos).

El término percepción se utiliza para indicar que el agente puede recibir entradas en cualquier instante. La secuencia de percepciones refleja el historial completo de lo que el agente ha recibido. Este tomará una decisión en un momento dado dependiendo de la secuencia completa de percepciones hasta ese instante.

En términos matemáticos se puede decir que el comportamiento del agente viene dado por la función del agente que proyecta una percepción dada en una acción. Esta función se puede presentar en forma de tabla.

Inicialmente, la función del agente para un agente artificial se implementará mediante el programa del agente. La primera es una descripción matemática abstracta, mientras que la segunda es una implementación completa, que se ejecuta sobre la arquitectura del agente.

Concepto de racionalidad

Un agente racional es aquel que hace lo correcto. Se puede decir que lo correcto es aquello que permite al agente obtener un resultado mejor. Por tanto, se necesita determinar una forma de medir el éxito, junto a la descripción del entorno y de los sensores y actuadores del agente.

***Medidas de rendimiento***

Las medidas de rendimiento incluyen los criterios que determinan el éxito en el comportamiento del agente. Cuando se sitúa un agente en un medio, éste genera una secuencia de acciones de acuerdo con las percepciones que recibe, esta secuencia hace que su hábitat pase por una secuencia de estados. Si la secuencia es la deseada, entonces el agente habrá actuado correctamente.

Como regla general, es mejor diseñar medidas de utilidad de acuerdo con lo que se quiere para el entorno, más que de acuerdo con cómo se cree que el agente debe comportarse.

***Racionalidad***

La racionalidad en un momento determinado depende de cuatro factores:

• La medida de rendimiento que define el criterio de éxito.

• El conocimiento del medio en el que habita acumulado por el agente.

• Las acciones que el agente puede llevar a cabo.

• La secuencia de percepciones del agente hasta este momento

Esto nos lleva a la definición de agente racional: En cada posible secuencia de percepciones, un agente racional deberá emprender aquella acción que supuestamente maximice su medida de rendimiento, basándose en las evidencias aportadas por la secuencia de percepciones y en el conocimiento que el agente mantiene almacenado.

***Omnisciencia, aprendizaje y autonomía***

Es necesario distinguir entre racionalidad y omnisciencia. Un agente omnisciente conoce el resultado de su acción y actúa de acuerdo con él, pero esto no es posible. Un ejemplo es si quiero cruzar la calle y cae una puerta de un avión cuando estoy cruzando, cruzar no fue irracional. La racionalidad no es lo mismo que la perfección. La racionalidad maximiza el rendimiento esperado, mientras la perfección maximiza el resultado real.

El asunto es que resulta imposible diseñar un agente que siempre lleve a cabo, de forma sucesiva, las mejores acciones después de un acontecimiento. La definición propuesta de racionalidad no requiere omnisciencia, ya que la elección racional depende sólo de la secuencia de percepción hasta la fecha.

Llevar a cabo acciones con la intención de modificar percepciones futuras, proceso denominado recopilación de información, es una parte importante de la racionalidad. Por ejemplo, <mirar> ambos lados antes de cruzar la calle. Un segundo ejemplo de recopilación de información lo proporciona la exploración que debe llevar a cabo el agente aspiradora en un medio inicialmente desconocido. Esto implica que el agente racional no sólo recopile información, sino que aprenda lo máximo posible de lo que está percibiendo.

La configuración inicial del agente puede reflejar un conocimiento preliminar del entorno, pero a medida que el agente adquiere experiencia éste puede modificarse y aumentar. Hay casos excepcionales en los que se conoce totalmente el entorno a priori, en estos casos no percibe ni aprende, simplemente actúa de forma correcta, pero son frágiles a cambios en su entorno.

Los agentes con éxito dividen las tareas de calcular la función del agente en tres períodos:

* Cuando se está diseñando el agente, los diseñadores están encargados de realizar algunos de estos cálculos
* Cuando está pensando en la siguiente operación, el agente realiza más cálculos
* cuando está aprendiendo de la experiencia, el agente lleva a cabo más cálculos para decidir cómo modificar su forma de comportarse

Un agente carece de autonomía cuando se apoya más en el conocimiento inicial que en sus propias percepciones. Un agente racional debe ser autónomo, debe saber aprender a determinar cómo tiene que compensar el conocimiento incompleto o parcial inicial. Por ejemplo, que la aspiradora aprenda donde y cuando aparecerá suciedad adicional.

Pocas veces se necesita autonomía completa desde el comienzo, pero es razonable proporcionar a los agentes que disponen de inteligencia artificial un conocimiento inicial, así como de la capacidad de aprendizaje. Después de las suficientes experiencias interaccionando con el entorno, el comportamiento del agente racional será efectivamente independiente del conocimiento que poseía inicialmente.

La naturaleza del entorno

Los entornos de trabajo son los «problemas» para los que los agentes racionales son las «soluciones»

***Especificación del entorno de trabajo***

En el diseño de un agente, el primer paso debe ser siempre especificar el entorno de trabajo de la forma más completa posible.

***Propiedades de los entornos de trabajo***

Se puede identificar un pequeño número de dimensiones en las que categorizar estos entornos. Estas dimensiones determinan, hasta cierto punto, el diseño más adecuado para el agente y la utilización de cada una de las familias principales de técnicas en la implementación del agente.

* **Totalmente observable vs Parcialmente observable**

Si los sensores del agente le proporcionan acceso al estado completo del medio en cada momento es **totalmente observable**

* **Determinista vs Estocástico**

Si el siguiente estado del medio está totalmente determinado por el estado actual y la acción ejecutada por el agente es **determinista**

* **Episódico vs Secuencial**

La experiencia del agente se divide en episodios que consisten en la percepción del agente y la realización de una única acción posterior. Si la elección de la acción en cada episodio depende sólo del episodio en sí mismo es **episódico**

* **Dinámico vs Estático**

Si el entorno puede cambiar cuando el agente está deliberando, entonces se dice que el entorno es **dinámico**

* **Discreto vs Continuo**

La distinción entre discreto y continuo se puede aplicar al estado del medio, a la forma en la que se maneja el tiempo y a las percepciones y acciones del agente. Por ejemplo, un medio con estados **discretos** como el del juego del ajedrez tiene un número finito de estados distintos. Por otro lado, el taxista define un estado **continuo**.

* **Agente individual vs Multiagente**

Un agente resolviendo un crucigrama por sí mismo está claramente en un entorno de agente **individual**, mientras que un agente que juega al ajedrez está en un entorno con dos agentes. Hay que analizar que entidades deben considerarse como agentes, la clave está en identificar si el comportamiento de B está mejor descrito por la maximización de una medida de rendimiento cuyo valor depende del comportamiento de A. En el ajedrez, la entidad oponente B intenta maximizar su medida de rendimiento, la cual, minimiza la medida de rendimiento del agente A, por tanto, es un entorno **multiagente competitivo**. En el medio del taxista, el evitar colisiones maximiza la medida de rendimiento de todos los agentes, así pues es un entorno **multiagente parcialmente cooperativo**.

Estructura de los agentes

El trabajo de la IA es diseñar el programa del agente que implemente la función del agente que proyecta las percepciones en las acciones, éste se ejecutará en algún tipo de computador con sensores físicos y actuadores, lo cual se conoce como arquitectura:

Agente = arquitectura + programa

La arquitectura hace que las percepciones de los sensores estén disponibles para el programa, ejecuta los programas, y se encarga de que los actuadores pongan en marcha las acciones generadas.

***Programas de los agentes***

Los programas de los agentes reciben las percepciones actuales como entradas de los sensores y devuelven una acción a los actuadores. Hay que tener en cuenta la diferencia entre los programas de los agentes, que toman la percepción actual como entrada, y la función del agente, que recibe la percepción histórica completa.

Agente dirigido mediante tabla: almacena la secuencia de percepciones y después las compara con las secuencias almacenadas en la tabla de acciones para decidir qué hacer. La tabla representa explícitamente la función que define el programa del agente. Para construir un agente racional de esta forma, los diseñadores deben realizar una tabla que contenga las acciones apropiadas para cada secuencia posible de percepciones.

Cuatro tipos básicos de programas para agentes:

* + Agentes reactivos simples
  + Agentes reactivos basados en modelos
  + Agentes basados en objetivos
  + Agentes basados en utilidad

***Agentes reactivos simples***

El tipo de agente más sencillo es el agente reactivo simple. Estos agentes seleccionan las acciones sobre la base de las percepciones actuales, ignorando el resto de las percepciones históricas. Se basa en conexiones establecidas en el programa del agente para que se ejecute una acción. Esta conexión se denomina regla de condición-acción, y se representa por:

sí <percepción> entonces <acción>

Texto, Carta

Descripción generada automáticamente

Una aproximación más general y flexible es la de construir primero un intérprete de propósito general para reglas de condición-acción y después crear conjuntos de reglas para entornos de trabajo específicos

Diagrama

Descripción generada automáticamente

La función INTERPRETAR-ENTRADA genera una descripción abstracta del estado actual a partir de la percepción, y la función REGLA-COINCIDENCIA devuelve la primera regla del conjunto de reglas que coincide con la descripción del estado dada. Hay que tener en cuenta que la descripción en términos de «reglas» y «coincidencias» es puramente conceptual, las implementaciones reales pueden otras.

Los bucles infinitos son a menudo inevitables para los agentes reactivos simples que operan en algunos entornos parcialmente observables. Salir de los bucles infinitos es posible si los agentes pueden seleccionar sus acciones aleatoriamente. Por ejemplo, si un agente aspiradora percibe [Limpio], puede lanzar una moneda y elegir entre Izquierda y Derecha. Este comportamiento puede resultar racional en algunos entornos multiagente pero no en entornos de agentes individuales.

***Agentes reactivos basados en modelos***

Para manejar la visibilidad parcial el agente debe mantener algún tipo de estado interno que dependa de la historia percibida y que de ese modo refleje por lo menos alguno de los aspectos no observables del estado actual. Primero, se necesita alguna información acerca de cómo evoluciona el mundo independientemente del agente. Segundo, se necesita más información sobre cómo afectan al mundo las acciones del agente.

Este conocimiento acerca de «cómo funciona el mundo», tanto si está implementado con un circuito booleano simple o con teorías científicas completas, se denomina modelo del mundo. Un agente que utilice este modelo es un agente basado en modelos.

Diagrama

Descripción generada automáticamente

Texto, Carta

Descripción generada automáticamente

Un agente reactivo basado en modelos, que almacena información sobre el estado actual del mundo utilizando un modelo interno. Después selecciona una acción de la misma forma que el agente reactivo.

***Agentes basados en objetivos***

El conocimiento sobre el estado actual del mundo no es siempre suficiente, por ejemplo, en un cruce, el taxista puede girar a la izquierda, derecha o seguir pero esto depende de a donde quiere ir. Además de la descripción del estado actual, el agente necesita algún tipo de información sobre su meta que describa las situaciones que son deseables.

En algunas ocasiones, la selección de acciones basadas en objetivos es directa, cuando alcanzar los objetivos es el resultado inmediato de una acción individual. En otras ocasiones, puede ser más complicado, cuando el agente tiene que considerar secuencias complejas para encontrar el camino que le permita alcanzar el objetivo

Diagrama

Descripción generada automáticamente

La toma de decisiones de este tipo es diferente de las reglas de condición–acción descritas anteriormente ya que hay que tener en cuenta consideraciones sobre el futuro.

Aunque el agente basado en objetivos pueda parecer menos eficiente, es más flexible ya que el conocimiento que soporta su decisión está representado explícitamente y puede modificarse.

***Agentes basados en utilidad***

Las metas por sí solas no son realmente suficientes para generar comportamiento de gran calidad en la mayoría de los entornos. El taxi puede llegar a destino por varias rutas pero algunas pueden ser más rápidas, seguras, fiables o baratas que otras. Para indicar que se prefiere un estado del mundo a otro es que un estado tiene más utilidad que otro para el agente.

Una función de utilidad proyecta un estado (o una secuencia de estados) en un número real, que representa un nivel de felicidad. Cuando hay objetivos en conflicto, por ejemplo velocidad y seguridad, la utilidad determina un equilibrio para cumplir ambos. Cuando hay varios objetivos y el cumplimiento de ellos es incierto, la utilidad pondera la probabilidad de éxito en función de la importancia de los objetivos.

Diagrama

Descripción generada automáticamente

Un agente basado en utilidad y basado en modelos. Utiliza un modelo del mundo, junto con una función de utilidad que calcula sus preferencias entre los estados del mundo. Después selecciona la acción que le lleve a alcanzar la mayor utilidad esperada, que se calcula haciendo la media de todos los estados resultantes posibles, ponderado con la probabilidad del resultado.

***Agentes que aprenden***

Turing (1950) consideró la idea de programar sus máquinas inteligentes a mano. El método que propone es construir máquinas que aprendan y después enseñarlas. El aprendizaje tiene otras ventajas, como se ha explicado anteriormente: permite que el agente opere en medios inicialmente desconocidos y que sea más competente que si sólo utilizase un conocimiento inicial.

Un agente que aprende se puede dividir en cuatro componentes conceptuales, tal y como se muestra en la figura.

Diagrama

Descripción generada automáticamente

La distinción más importante entre el **elemento de aprendizaje** y el **elemento de actuación** es que el primero está responsabilizado de hacer mejoras y el segundo se responsabiliza de la selección de acciones externas.

El elemento de aprendizaje se realimenta con las críticas sobre la actuación del agente con respecto a un nivel de actuación fijo y determina cómo se debe modificar el elemento de actuación para proporcionar mejores resultados en el futuro. Dado un diseño para un agente, se pueden construir los mecanismos de aprendizaje necesarios para mejorar cada una de las partes del agente.

El último componente del agente con capacidad de aprendizaje es el **generador de problemas**. El agente puede llevar a cabo nuevas acciones que no sean totalmente óptimas a corto plazo, pero puede descubrir acciones mejores a largo plazo. El generador de problemas es resp