

Inteligencia Artificial Un enfoque moderno

Cap 1 Introducción

¿Qué es la IA?

Se habla de la IA como algo fascinante, pero para esto necesitamos saber ¿Qué es la IA?, de las cuatro definiciones que aparecen en el libro, indican que la ideal para el concepto de inteligencia es la racionalidad, refiriéndose a que un sistema es racional cuando hace lo correcto con respecto a lo que conoce.

Para entender mejor los conceptos o las referencias que aluden a estos conceptos sobre Sistemas que piensan como humanos, Sistemas que piensan racionalmente, Sistemas que actúan como humanos y Sistemas que actúan racionalmente, los explicaré de una manera más simple.

Comportamiento Humano: El enfoque de la Prueba de Turing

La prueba consiste en que un evaluador humano analice las respuestas, de una serie de preguntas y de las cuales son contestadas por un humano y cuales por una entidad inteligente, si el evaluador no era capaz de diferenciar las respuestas, la entidad supera la prueba, se menciona que para ese entonces no se podría superar esa prueba, que el computador necesitara las siguientes capacidades:

- Procesamiento de Lenguaje Natural: para que pueda comunicarse en el idioma deseado
- Representación del Conocimiento: que pueda almacenar conocimiento o lo que siente
- Razonamiento Automático: para poder hacer uso de el conocimiento almacenado que tiene y responder preguntas y obtener nuevas conclusiones
- Aprendizaje Automático: para que pueda adaptarse a cambios y detectar patrones

Pero en la actualidad, el primer pensamiento que se me vino a la mente es que las herramientas como bing, chat-gpt, Bard, etc. ya pueden hacer eso y hasta más. Pero la prueba de Turing se limitaba y no tenía interacción física para evaluar al computador, que para medir la inteligencia es necesario simular también físicamente a una persona. La Prueba Global de Turing agrega que para que pueda pasar la prueba física debe de pasar objetos físicos por una ventana y para ello necesitaba de:

- Visión computacional: que le permita reconocer objetos:
- Robótica: para poder manipular y mover objetos.

En este punto se me vinieron a la mente los robots de Boston Dynamics, que recientemente se ve como realizan movimientos en distintas pruebas de equilibrio y superar obstáculos, donde empezaron desde que pudieran tener equilibrio, luego aprender a caminar, correr y superar obstáculos. Lo que se habla es que se deje de tratar de hacer parecer o duplicar dicho ejemplar y entender más los fundamentos de lo que hace dicho ejemplar, como mencionan con los hermanos Wright, que entendieron que tenían que dejar de simular a los pájaros y entender el por qué? y como? es que vuelan los pájaros, comprendiendo los principios de la aerodinámica.

Pensar como un Humano: El enfoque del Modelo Cognitivo

En la teoría que se propone dice que si los datos de entrada, salida y el tiempo de reacción que toma para un computador es el mismo que de un humano, se puede decir que en efecto una máquina puede pensar como un humano, pero realmente alguna vez las máquinas pueden o podrían llegar a pensar como humanos si estas usan modelos matemáticos para “Simular” y cuando aún no se conoce realmente como funciona o cómo es que ocurre el funcionamiento de las mentes humanas?. Pero es cierto que hay similitudes y diferencias entre las técnicas de IA y cognición humana. Aquí es donde menciona el libro la convergencia de la ciencia cognitiva y los modelos de IA y cómo esto ha permitido el avance paralelo de estas disciplinas apoyándose mutuamente.

Pensamiento Racional: El enfoque de las Leyes del Pensamiento

¿Cual es la manera correcta de pensar?, bueno pues cual es la manera más común que decimos que pensamos, lógicamente no?, como en el libro, que menciona los silogismos, donde si tenemos la premisas correctas, podemos sacar el resultado correcto, ponen de ejemplo: “Sócrates es un Hombre, todos los hombres son mortales, por lo tanto Sócrates es mortal”. Pero no es fácil pasarle a los sistemas la notación adecuada de la lógica, si está el conocimiento de manera informal y queremos pasarlo a la llamada tradición lógica y porque no es lo mismo resolver un problema en principio y hacerlo ya en práctica.

Actuar de Forma Racional: el enfoque del agente Racional

Un Agente es Racional, cuando actúa con racionalidad para alcanzar el mejor resultado y si hay desconocimiento o incertidumbre, es obtener el mejor resultado esperado. En el enfoque de la IA con las leyes del pensamiento, se enfoca en hacer las inferencias correctas y tener las inferencias correctas puede hacer que se considere un agente racional, aunque a veces no se necesita la racionalidad para tomar decisión, en ocasiones solo se tiene que tomar una decisión y ya y también hay formas de actuar racionalmente sin hacer inferencias, ahora bien para poder tomar una decisión es bueno tener una percepción visual del entorno para poder tener un contexto más nutrido y poder llegar a tomar la decisión que más se adapte a las circunstancias.

Los fundamentos de la Inteligencia Artificial

Veremos las distintas disciplinas que han aportado conocimientos y contribuido ideas, técnicas y puntos de vista para el desarrollo de la IA. También veremos un poco de historia sobre cómo estas influyeron en la IA al paso de los años, con el fin de hacer avanzar a la IA.

Filosofía

- ¿Se pueden utilizar reglas formales para extraer conclusiones válidas?
- ¿Cómo se genera la inteligencia mental a partir de un cerebro físico?
- ¿De dónde viene el conocimiento?
- ¿Cómo se pasa del conocimiento a la acción

Aristoteles fue de los primeros en generar un conjunto de leyes que rigieron la parte racional de la inteligencia. Desarrolló un sistema para razonar de la mejor forma, usando silogismos, con los cuales se podían sacar conclusiones partiendo de premisas. Después Ramón Dull tuvo la idea de que se podía adquirir el razonamiento útil de manera artificial. Thomas Hobbes propuso que razonar era como la computación numérica, en el sentido de que nosotros sumamos y restamos silenciosamente en nuestros pensamientos. Blaise Pascal

escribió que la máquina aritmética hacía más efectos parecidos a los pensamientos, que a las acciones animales. Gottfried Wilhelm Leibniz hizo un dispositivo mecánico con el fin de realizar operaciones sobre conceptos en lugar de números, pero el campo de acción fue limitado.

Sabiendo que una serie de reglas pueden describir la parte racional y formal de la mente, sigue entender la mente como un sistema físico. René Descartes puso en discusión la diferencia entre la mente, la materia y los problemas que surgen. Pues si se basa en que la concepción de la mente la vemos puramente física, no habría espacio para el libre albedrío, ya que el libre albedrío es fundamental a la hora de tomar decisiones, cada quien puede tomar una decisión siendo influenciada por el conocimiento aplicando el racionalismo, pero Descartes defensor del dualismo dice que la mente se ve influenciada por la naturaleza y las leyes físicas y que a diferencia de los animales, estos no tienen ese dualismo y de cierta manera considerarlos máquinas, la alternativa al dualismo es el materialismo que dice que las operaciones del cerebro se realizan de acuerdo a que las leyes de la física constituyen la mente. Entonces el libre albedrío es la forma en la que la percepción de las opciones disponibles aparecen en el proceso de selección.

Dado que una mente física maneja el conocimiento, ahora sigue encontrar la fuente de este conocimiento. El movimiento empírico iniciado por Novum Organum, de Francis Bacon, se representaba por la frase que decía: "Nada existe en la mente que no haya pasado antes por los sentidos". David Hume propuso lo que ahora se conoce como el principio de inducción a las reglas generales se adquieren por medio de la exposición repetidas de las mismas acciones o hechos. El famoso círculo de Viena desarrolló la doctrina de positivismo, la cual argumenta que el conocimiento se puede caracterizar por medio de teorías lógicas, relacionadas con la observación que produce estímulos sensoriales. Carnap y Carl Hempel tenían la teoría de la confirmación, que explicaba que el conocimiento se adquiere a partir de la experiencia. El libro de Carnap "The logical Structure of the world", define un procedimiento computacional para poder obtener conocimiento a partir de experiencias primarias. Fue posiblemente una de las primeras teorías en mostrar la mente como un proceso computacional.

Otra discusión filosófica de la mente, es la relación que hay entre conocimiento y acción. Para la IA esta parte es esencial, ya que la inteligencia necesita de acción como razonamiento. Ya que, con el hecho de entender cómo se justifican ciertas acciones, se puede llegar a saber como crear un agente que pueda tomar acciones y estas sean justificables o racionales. Aristoteles argumentaba que las acciones son justificables por la conexión lógica que hay entre el objetivo y el conocimiento del efecto de dichas acciones. Aristoteles siguió trabajando en este temas, donde reflexiona sobre los medios y no el fin en si, habla de que si se puede llegar al fin, hay que evaluar los medios y si por un medio no se puede, pues ver por que no e intentar con otro hasta llegar al fin y si no hay un medio se llega a un estado imposible y se abandona la búsqueda peor si hay posibilidad siempre se va a intentar llegar a la causa.

Matemáticas

- ¿Qué reglas formales son las adecuadas para obtener conclusiones válidas?
- ¿Qué se puede computar?
- ¿Cómo razonamos con información incierta?

Para pasar de la filosofía, que puso los límites más importantes de la IA a una ciencia un poco más formal, se necesita una formulación matemática en tres apartados fundamentales como lógica, computación y probabilidad.

El concepto de lógica se basa en los filósofos de la Grecia antigua, pero el desarrollo matemático de esta comenzó con George Boole que definió la lógica Booleana. Gottlob Frege añadió a la lógica de Boole los objetos y relaciones, creando la lógica de primer orden que se usa hoy día como la representación de conocimiento más básica. Luego Alfred Tarski con la teoría de referencia enseñó cómo relacionar objetos de una lógica con objetos de la realidad. El próximo paso era delimitar lo que se podía hacer con la lógica y la informática.

Se cree que el primer algoritmo no trivial fue el Euclídeo para calcular el máximo común divisor. El considerar los algoritmos como objetos se remonta a la época de al-Khowarazmi, un matemático persa que con sus escritos introdujo los números arábigos y el álgebra a Europa. Boole presentó algoritmos para hacer inferencias lógicas, al final del siglo XIX se formalizó el razonamiento matemático general con la lógica deductiva. David Hilbert presentó problemas a los cuales los matemáticos de esa época se iban a enfrentar, uno de esos problemas era de que si hay un algoritmo que pudiera comprobar la validez de una proposición lógica en la que se presenten números naturales. Kurt Gödel demostró que si existe un procedimiento eficaz para demostrar cualquier afirmación que sea verdadera en la lógica de primer orden, pero con la lógica de primer orden no se podía capturar el principio de inducción matemática que se necesitaba para caracterización de los números naturales, también mediante el teorema de incompletitud demostró en cualquier lenguaje que tuviera la capacidad para expresar las propiedades de los números naturales, hay aseveraciones verdaderas que no se pueden establecer, en el sentido de verificar su validez con ningún algoritmo.

Lo anterior se puede tomar como que los números enteros tienen alguna función que no son posibles de representar mediante un algoritmo, es decir que no se puedan calcular. Esto llevó a Alan Turing a tratar las características de las funciones que sí se pueden representar, esto suele presentar problemas en cierto punto, porque no es posible dar una definición formal de cálculo o procedimiento eficaz. La Tesis de Church-Turing, afirma que la máquina de Turing puede calcular cualquier función computable, goza de aceptación generalizada ya que da una definición suficiente. Turing además demostró que existen algunas funciones que no son posibles de calcular con la máquina de Turing. Por ejemplo, ninguna máquina puede decidir si un programa dado dará una respuesta a partir de la entrada o si seguirá calculando indefinidamente.

La Intratabilidad tuvo sus influencias importantes. Se dice que un problema es intratable si el tiempo que toma resolver el problema empieza a crecer exponencialmente. La diferencia entre crecimiento exponencial y polinomial. Un crecimiento exponencial conlleva a que un problema pueda ser imposible de resolver en un tiempo razonable, por eso es importante optar por dividir la problemática en pequeñas partes para que sean más óptimos de resolver, en vez de uno intratable.

¿Cómo se puede saber si un problema es intratable?, la teoría NP-completitud propone una forma. Cook y Karp demostraron que hay tipos de problemas de razonamiento y búsqueda

combinatoria que son NP completos. Todo tipo de problemas que se puedan clasificar como NP completos, se podría decir que son intratables aunque no se ha probado que necesariamente un problema np sea intratable. Los resultados de esto, fueron contrastados por cómo la gente recibió a los primeros computadores “Supercerebros” o “¡Más rápidos que Einstein!” y aunque eran muy veloces, se distinguían más por su manejo de los recursos. La IA ha ayudado recientemente a la confirmación o desacreditación de los problemas np, diciendo los problemas np completos son complicados o sencillos de resolver.

La tercera mayor aportación de las matemáticas a la IA fue la teoría de la probabilidad. El primero en ponerla sobre la mesa fue el italiano Gerolamo Cardano, la presentó con referencias a juegos de apuestas. La probabilidad se convirtió en una pieza fundamental para la ciencia cuantitativa, en mediciones con incertidumbre y teorías incompletas. Fermat, Pascal, Bernoulli y Laplace, entre otros, ayudaron a que avanzara cuando introdujeron los cálculos estadísticos. Thomas Bayes propuso una regla para las probabilidades que necesitaban una actualización debido a nuevos datos. La regla de Bayes y el análisis Bayesiano son parte de las bases que forman las propuestas modernas para manejar el razonamiento incierto en sistemas de la IA.

Economía

- ¿Cómo se debe llevar a cabo el proceso de toma de decisiones para maximizar el rendimiento?
- ¿Cómo se deben llevar a cabo acciones cuando otros no colaboren?
- ¿Cómo se deben llevar a cabo acciones cuando los resultados se obtienen en un futuro lejano?

A pesar de que los antiguos griegos hicieron contribuciones al pensamiento económico, tal cual, la ciencia de la economía empezó con el filósofo escocés Adam Smith, pues fue de los primeros en tratar la economía como una ciencia, usando la idea de donde las economías pueden tratarse como un conjunto de agentes individuales que buscan maximizar su bienestar económico. La gran mayoría cree que la economía solo se trata del dinero, pero los economistas dicen que lo que en verdad estudian son los comportamientos de las personas, las decisiones que toman para obtener un beneficio que desean o esperan. Léon Walras formalizó el tratamiento matemático del “beneficio esperado” o “utilidad”.

La teoría de la decisión, combina la teoría de la probabilidad y la utilidad, proporcionando una visión más completa y formal sobre la toma de decisiones ya sean económicas o de otra disciplina, que se realizan con un grado de incertidumbre, esto, en casos donde se capturan de manera correcta las descripciones probabilísticas en función en la forma en la que se toman las decisiones en el entorno; Esto es adecuado para grandes economías, en las que un agente individual no depende de las acciones de los demás agentes. Para casos donde las pequeñas economías, las situaciones son similares a las de un juego, donde las acciones de un individuo pueden afectar directamente la utilidad de otro para bien o para mal. Los desarrollos de Von Neuman y Morgenstern desde la teoría de juegos señalaba el hecho de que un agente racional debía de actuar aleatoriamente, al menos en función de los otros jugadores. Los economistas no se preocupaban por la tercera cuestión, que era como actuar de manera racional cuando no se esperaban un resultado de inmediato, cuando el resultado de las acciones se obtienen de manera secuencial. El área de la investigación operativa está detrás de este objetivo, esto surgió debido a la necesidad del

Reino Unido para mejorar las instalaciones de radares en la Segunda Guerra Mundial y después en la toma de decisiones civiles complejas. Richard Bellman en su trabajo formalizó una serie de problemas de decisión secuencial llamados proceso de decisión de Markov .

Los trabajos en Investigación Operativa y Economía han ayudado en gran parte a la noción de agente racional que se presenta, aunque en el campo de investigación de la IA se desarrollaba por partes distintas. Herbert Simon quien ganó el premio nobel de economía, fue de los primeros en investigar en el campo de la IA, dice en su trabajo que los modelos basados en satisfacción daban una descripción sobre el comportamiento humano real.

Neurociencia

- ¿Cómo procesa la información el cerebro?

La neurociencia estudia el sistema neurológico y en especial el cerebro. Si bien no se sabe exactamente cómo el cerebro forma el conocimiento/pensamiento siendo este el más grande misterio de la ciencia, se sabe que este está relacionado con el proceso del pensamiento, puesto que fuertes contusiones pueden llegar a afectar y causar cierta incapacidad mental. El cerebro humano es diferente en ciertos aspectos en comparación a otros seres vivos, si comparamos el tamaño se puede decir que es el más grande en proporción al tamaño. Fue hasta mediados del siglo XVIII que se llegó a la conclusión que efectivamente el cerebro es la base de la conciencia.

Paul Broca llevó a cabo un estudio sobre la dificultad del habla (afasia) donde estudió a pacientes con daño cerebral, donde convenció a la sociedad que el cerebro también tenía funciones cognitivas específicas. Mostró que el habla o la función del habla se llevaba a cabo en el hemisferio izquierdo conocida hoy en día como el área de Broca. En aquel entonces ya se sabía que el cerebro estaba compuesto por células nerviosas/neuronas, no fue hasta el año 1873 que Camillo Golgi elaboró una técnica de coloración para observar individualmente las neuronas del cerebro.

Actualmente se conoce sobre la relación que existe entre las áreas del cerebro y las partes del cuerpo que controla o de las que recibe impulsos sensoriales. Estas relaciones pueden cambiar espontáneamente en pocas semanas, algunos animales pueden disponer de una variedad de posibilidades. Sin embargo aún no está del todo claro el cómo algunas áreas se pueden encargar de funciones que eran hechas por áreas dañadas. Tampoco hay algo que nos describa o explique cómo se almacenan los recuerdos individuales. El estudio de la actividad en cerebros intactos empezó en 1929 con el descubrimiento del electroencefalograma que desarrolló Hans Berger. El reciente descubrimiento de las imágenes de resonancia magnética funcional, le está dando a los neurólogos imágenes detalladas de la actividad cerebral, permitiendo obtener medidas que corresponden con procesos cognitivos. Este campo ha evolucionado en parte a los avances en los estudios en celdas individuales y la actividad neuronal, a pesar de los avances, todavía queda un gran camino para llegar a entender el funcionamiento de los procesos cognitivos. Lo impactante es que un conjunto de células puedan llegar a generar conocimiento, acción y conciencia o lo que se conoce como inteligencia.

Si bien tratamos de asimilar el funcionamiento del cerebro en las computadoras, estos tienen propiedades distintas en casos de tareas distintas, hay 1000 veces más neuronas en un

cerebro que compuertas lógicas en la CPU de un computador estándar. La ley de Moore predice que el número de compuertas, se igualará al número de neuronas, alrededor del año 2020. También la diferencia en la capacidades de almacenamiento no se pueden comparar con las diferencias de procesamiento y paralelismo. Los circuitos procesan una instrucción en nanosegundos, las neuronas son millones de veces más lentas. Las neuronas y la sinapsis del cerebro están activas simultáneamente, mientras que los computadores actuales tienen una o varias CPU, incluso sabiendo que una computadora puede ser millones de veces más rápido en la velocidad de intercambio, el cerebro acaba siendo más rápido en lo que hace.

Psicología

- ¿Cómo piensan y actúan los humanos y los animales?

La Psicología científica empezó con Hermann von Helmholtz y su discípulo Wundt, Wundt abrió el primer laboratorio de psicología experimental en Leipzig. Puso mucho empeño en realizar experimentos controlados, en donde los operadores hacían tarea de percepción o asociación al tiempo que sometían a introspección sus procesos mentales. Los meticulosos controles evolucionaron hasta que convirtieron la psicología en una ciencia, pero la misma naturaleza de los datos hizo que fuera poco probable que un investigador pudiera contradecir sus propias teorías. Los Biólogos también estudiaron el comportamiento donde los datos no eran introspectivos y se desarrolló una metodología objetiva. El movimiento conductista aplicó este punto de vista a los humanos, rechazando cualquier teoría en la que procesos mentales se vieran involucrados, con el argumento de que la introspección no provee evidencia confiable. Los conductistas insistieron en estudiar exclusivamente mediciones objetivas de percepciones y/o estímulos sobre animales y las acciones que resultaban como respuestas de los estímulos. Construcciones mentales como conocimientos, creencias, adjetivos y pasos en un razonamiento se consideran más como psicología popular y no científica. El conductismo hizo muchos descubrimientos experimentando con ratas y palomas, pero con humanos tuvo menor éxito en la comprensión de estos.

La Psicología cognitiva pone énfasis en la percepción y esta converge con cierto tipo de inferencia lógica inconsciente. Este punto de vista se vio opacado por la psicología conductista pero en la universidad de Cambridge los modelos cognitivos tomaron fuerza. La obra de Kenneth Craik reestablece que los términos mentales como creencias y objetivos son igual de válidos científicamente como lo es la presión y la temperatura cuando se refiere a gases sin importar que estén formados por moléculas que no tienen presión y gases. Craik establece tres elementos clave para diseñar un agente en base a conocimiento:

1. El estímulo debe de ser traducido a una representación interna
2. La representación se debe manipular mediante procesos cognitivos para generar representaciones internas
3. Está a su vez se traducen de nuevo en acciones

Argumentando que si un organismo tiene un modelo de la realidad y todas las posibilidades, podrá tomar la mejor decisiones a partir de distintas opciones, decidir cuál es la más óptima en base a su conocimiento.

La ciencia cognitiva se puede decir que empezó en el MIT. George Miller presentó tres artículos, que influyeron en cómo se podrían utilizar los modelos informáticos para modelar

la psicología de la memoria, lenguaje y el pensamiento lógico. Actualmente los psicólogos tienen un punto de vista donde la teoría cognitiva debe ser como un programa computacional, en pocas palabras se debe describir un mecanismo de procesamiento de información detallado para la implementación de algunas funciones cognitivas.

Ingeniería Computacional

- ¿Cómo se puede construir un computador eficiente?

Se necesitan dos cosas para que la IA pueda materializarse, esas cosas son inteligencia y un artefacto, este artefacto es el computador. El computador electrónico digital se hizo de manera independiente y simultáneamente por distintos científicos en la segunda guerra mundial. Alan Turing y su equipo hicieron el primero operacional electromecánico, con un único propósito: descifrar los mensajes encriptados de los alemanes. Luego Colossus en 1943, está de propósito general basada en válvulas de vacío. Konrad Zuse inventó el primer computador operacional programable, Zuse también inventó los números de coma flotante y el primer lenguaje de programación de alto nivel, Plankalkül. El ABC, el primer computador electrónico creado por Atanasoff y Berry. El ENIAC es un desarrollo en el ámbito militar secreto, en la Universidad de Pensilvania por un grupo donde destacan Mauchl y Eckert, se considera como el precursor de los computadores modernos.

Desde mediados del siglo anterior, con cada generación de dispositivos hardware han conllevado a un aumento en la velocidad de proceso y en la capacidad de almacenamiento, así como la reducción del precio. Se dice que se duplicará la potencias de estos cada 1.8 años por las siguientes dos décadas y después de eso se necesitarán otras tecnologías y la ingeniería molecular para seguir mejorando.

Desde antes de que llegaran los computadores, ya existían dispositivos de cálculo. La primera máquina programable de el tellar, que usaba tarjetas perforadas para almacenar información sobre los patrones de los bordados, está desarrollada por Jacquard. Charles Babbage diseñó dos máquinas que no pudo desarrollar, la “Máquina de diferencias” que tenía como objetivo facilitar los cálculos de las tablas matemáticas para proyectos científicos y de ingeniería y la “Maquina Analítica” era más ambiciosa, incluía memoria direccionable, programas almacenados y saltos condicionales. Fue el primero en tener capacidades necesarias para llevar a cabo la computación universal. Ada Lovelace fue la primera programadora, el lenguaje ADA se llama así en honor a ella. Escribió programas para la máquina analítica e inclusive estuvo especulando acerca de la posibilidad de que la máquina jugara ajedrez y compusiera música.

La IA tenía una deuda con la parte de software que ha proporcionado los sistemas operativos, lenguajes de programación y las herramientas que han ayudado a hacer los programas modernos. En parte esta deuda se ha solventado, las investigaciones sobre la IA han dado ideas que han beneficiado a la informática, como el tiempo compartido, los computadores personales con interfaces gráficas y ratones., entornos de desarrollo rápidos, listas enlazadas, administración automática de memoria y conceptos clave de la programación simbólica, funcional, dinámica y orientada a objetos.

Teoría de Control y Cibernética

- ¿Cómo un artefacto puede operar bajo su mismo control?

La teoría de control y cibernética ha experimentado un desarrollo significativo desde la antigüedad hasta la actualidad. En sus inicios, Ktesibios de Alejandría construyó la primera máquina autocontrolada, un reloj de agua con un regulador que mantenía un flujo constante y predecible. A lo largo de la historia, destacan inventos como el motor de vapor de James Watt y el termostato de Cornelis Drebbel, que ejemplifican sistemas auto regulables y retroalimentados.

El matemático Norbert Wiener fue una figura central en el desarrollo de la teoría de control. Junto con colegas como Arturo Rosenblueth y Julian Bigelow, desafiaron la ortodoxia conductista al considerar el comportamiento determinista como emergente de un mecanismo regulador que busca minimizar el error. En la década de 1940, Wiener, Warren McCulloch, Walter Pitts y John von Neumann organizaron conferencias que exploraban modelos matemáticos y computacionales en cognición, influyendo en las ciencias del comportamiento.

Wiener, en su libro "Cybernetics" (1948), dio a conocer al público las posibilidades de las máquinas con inteligencia artificial. Aunque la teoría de control moderna, especialmente el control óptimo estocástico, comparte el objetivo de diseñar sistemas que maximizan una función objetivo en el tiempo, la inteligencia artificial (IA) y la teoría de control se separaron en gran medida debido a las diferencias en las técnicas matemáticas utilizadas y en los conjuntos de problemas abordados. La IA, surgida en los años 50, buscó superar las limitaciones matemáticas de la teoría de control utilizando herramientas de inferencia lógica y computación para abordar problemas relacionados con el lenguaje, la visión y la planificación, que estaban más allá del alcance de la teoría de control.

Lingüística

En 1957, B. F. Skinner publicó "Verbal Behavior", una obra que abordaba el aprendizaje del lenguaje desde la perspectiva conductista. Sin embargo, Noam Chomsky, en su obra "Syntactic Structures", criticó esta teoría al señalar que no explicaba la creatividad en el lenguaje, especialmente la capacidad de los niños para comprender y construir oraciones nunca antes escuchadas. La teoría de Chomsky, basada en modelos sintácticos que se remontaban a Panini (350 a.C.), ofrecía una explicación más completa y poseía el formalismo necesario para su programación.

La crítica de Chomsky marcó un cambio significativo en el paradigma lingüístico, alejándose del conductismo y dando lugar a la lingüística moderna y a la inteligencia artificial (IA). Ambos campos evolucionaron de manera interconectada, convergiendo en la lingüística computacional o procesamiento del lenguaje natural. A medida que avanzaba la investigación, se hizo evidente que el entendimiento del lenguaje era más complejo de lo inicialmente pensado en 1957. No se trataba solo de comprender la estructura de las oraciones, sino también de comprender la materia y el contexto subyacente. Esta comprensión más amplia del lenguaje se volvió crucial, y la lingüística computacional se desarrolló para abordar problemas de representación del conocimiento y búsqueda de información relacionados con el lenguaje. Este enfoque se nutrió de décadas de investigación en el análisis filosófico del lenguaje.

Cap 2 Agentes Inteligentes

La búsqueda de comprender y diseñar agentes inteligentes ha llevado a la formulación y exploración del concepto de racionalidad en diversos entornos y contextos. En este capítulo, se abordará cómo el principio de racionalidad puede aplicarse de manera versátil a una amplia gama de agentes, sin importar el medio en el que operen. La premisa fundamental es desarrollar un conjunto de principios de diseño que sirvan como cimiento para la construcción de agentes útiles, aquellos que se pueden considerar razonablemente inteligentes.

La exploración comienza con un análisis detenido de los agentes, sus entornos y las complejas interacciones que surgen entre ellos. Al observar que algunos agentes exhiben un desempeño superior a otros, surge naturalmente la noción de un "agente racional", definido como aquel que se comporta de la mejor manera posible dadas las circunstancias. La actuación de estos agentes está intrínsecamente vinculada a la naturaleza de su entorno, reconociendo que algunos hábitats presentan desafíos más complejos que otros.

En este contexto, se presenta una categorización inicial del medio, destacando cómo las propiedades específicas de un hábitat influyen en el diseño de agentes adaptados a esas condiciones particulares. Se esbozan varios "esquemas" básicos para el diseño de agentes, los cuales serán desarrollados y ampliados a lo largo de este estudio. Este capítulo sienta las bases para una comprensión más profunda de la racionalidad en el diseño de agentes inteligentes, explorando cómo los principios fundamentales pueden ser aplicados en diversos escenarios para lograr la creación de sistemas verdaderamente inteligentes y eficientes.

Agentes y su Entorno

En el ámbito de la inteligencia artificial, un agente se define como cualquier entidad que tiene la capacidad de percibir su entorno a través de sensores y de tomar acciones en ese entorno mediante actuadores. Este concepto abarca desde agentes humanos con órganos sensoriales y partes del cuerpo, hasta agentes robóticos que reciben información a través de teclados y redes para luego actuar mediante la emisión de mensajes y ejecución de comandos.

Se parte de la premisa de que cada agente puede percibir sus propias acciones, aunque no siempre sus efectos. La percepción se refiere a la capacidad del agente para recibir entradas en cualquier momento, y la secuencia de percepciones constituye el historial completo de lo que el agente ha experimentado. La toma de decisiones por parte del agente en un momento dado se basa en la secuencia completa de percepciones hasta ese instante. La especificación de las decisiones para todas las posibles secuencias de percepciones constituye una descripción completa del comportamiento del agente.

En términos matemáticos, el comportamiento del agente se expresa mediante una función que mapea una percepción dada en una acción correspondiente. Aunque esta función podría presentarse en forma de tabla, en muchos casos sería impracticable debido a su extensión (potencialmente infinita a menos que se limite la secuencia de percepciones considerada). La función del agente, una descripción abstracta y matemática, se distingue de su programa, que es la implementación concreta ejecutada sobre la arquitectura del agente.

En resumen, este enfoque analiza la relación entre la percepción y la acción de los agentes, proponiendo un marco conceptual que abarca desde agentes humanos hasta entidades artificiales, y destaca la distinción entre la función abstracta del agente y su implementación práctica a través del programa. Este análisis es esencial para comprender y diseñar agentes inteligentes en diversos contextos y aplicaciones.

Buen comportamiento: el concepto de racionalidad

En el contexto de la inteligencia artificial, la noción de un agente racional se define como aquel que realiza acciones correctas, y conceptualmente, cada entrada en la tabla que define su función debe ser completada de manera adecuada. La acción correcta se entiende como aquella que optimiza el rendimiento esperado del agente. El éxito se mide a través de las denominadas medidas de rendimiento, que determinan el comportamiento exitoso del agente en su entorno.

Las medidas de rendimiento reflejan la efectividad del agente al generar secuencias de acciones que llevan al entorno a través de diferentes estados. La elección de medidas de rendimiento objetivas es crucial, ya que la percepción subjetiva del agente puede no ser fiable. Se destaca la importancia de diseñar medidas de utilidad de acuerdo con los objetivos del entorno en lugar de basarse en expectativas sobre cómo debería comportarse el agente.

El concepto de racionalidad en un agente se compone de cuatro factores: la medida de rendimiento que define el éxito, el conocimiento acumulado por el agente sobre su entorno, las acciones que el agente puede emprender y la secuencia de percepciones hasta el momento. La definición resultante de un agente racional establece que en cada secuencia de percepciones posible, el agente debe emprender la acción que maximice su medida de rendimiento, basándose en las percepciones y el conocimiento almacenado.

Se discute la distinción entre racionalidad y omnisciencia, subrayando que la racionalidad busca maximizar el rendimiento esperado, mientras que la omnisciencia implica conocer el resultado real de las acciones. Además, se aborda la importancia del aprendizaje y la autonomía en los agentes racionales, indicando que la capacidad de aprender de la experiencia es esencial para la autonomía y el éxito sostenido en diversos entornos. La autonomía se define como la capacidad del agente para depender menos del conocimiento inicial proporcionado por el diseñador y más de sus propias percepciones y aprendizaje. En última instancia, se enfatiza que la incorporación del aprendizaje facilita el diseño de agentes racionales que pueden tener éxito en una variedad de contextos.

La naturaleza del Entorno

Describe la importancia de comprender la naturaleza del entorno al diseñar agentes racionales. Se destaca la especificación del entorno de trabajo (REAS: Rendimiento, Entorno, Actuadores, Sensores) como un paso fundamental en el diseño de agentes. Se utiliza el ejemplo de un taxista automático para ilustrar la complejidad de los entornos y cómo influyen en el diseño del programa del agente.

Se mencionan varias propiedades de los entornos, como totalmente observable vs. parcialmente observable, determinista vs. estocástico, episódico vs. secuencial, estático vs.

dinámico, discreto vs. continuo y agente individual vs. multiagente. Estas propiedades influyen en la complejidad y el diseño de los agentes.

Se destaca que muchos entornos del mundo real son parcialmente observables, estocásticos, secuenciales, dinámicos, continuos y multiagentes, lo que agrega complejidad al diseño de agentes. Se menciona que, a pesar de las clasificaciones iniciales, la complejidad de la realidad a menudo hace que los entornos se traten prácticamente como estocásticos.

El texto concluye mencionando que el repositorio de código asociado al libro incluye implementaciones de varios entornos y un simulador para evaluar agentes en diferentes condiciones. Se enfatiza la importancia de diseñar agentes que maximicen el rendimiento medio en una clase de entornos.

Estructura de los agentes

Se aborda la problemática del diseño de programas para agentes de inteligencia artificial (IA), que explora diferentes enfoques para implementar la función del agente que mapea percepciones a acciones. Se destaca la dificultad de diseñar programas efectivos, especialmente al considerar el enorme espacio de búsqueda en entornos complejos.

El autor introduce cuatro tipos básicos de programas para agentes:

1. ****Agentes Reactivos Simples:**** Toman decisiones basadas únicamente en percepciones actuales, ignorando la historia. Su simplicidad limita su inteligencia.
2. ****Agentes Reactivos Basados en Modelos:**** Mantienen un estado interno que refleja aspectos no observables del entorno. Utilizan modelos del mundo para tomar decisiones.
3. ****Agentes Basados en Objetivos:**** Consideran metas deseadas además del estado actual del mundo. Pueden planificar y buscar secuencias de acciones para alcanzar sus objetivos.
4. ****Agentes Basados en Utilidad:**** Introducen funciones de utilidad para evaluar la calidad de los estados del mundo. Buscan maximizar la utilidad en lugar de simplemente alcanzar metas.

El texto destaca la importancia de la aprendizaje en la IA y presenta un modelo conceptual de agentes que aprenden. Se enfatiza la retroalimentación entre el elemento de actuación, la crítica y el generador de problemas, así como la modificación continua de componentes para mejorar el rendimiento general del agente.

Se reconoce que el aprendizaje en la IA implica ajustar cada componente del agente para que se comporte de manera más coherente con la información recibida, mejorando así su rendimiento promedio.

Cap 26 Fundamentos Filosóficos:

En el desarrollo de la inteligencia artificial (IA), los filósofos han desempeñado un papel crucial mucho antes de que los computadores se convirtieran en una realidad tangible. Este capítulo se sumerge en los fundamentos filosóficos que han preexistido a la era de la IA, explorando las inquietudes fundamentales que los pensadores han abordado durante siglos. Desde la pregunta esencial sobre el funcionamiento de la mente hasta la indagación sobre la posibilidad de que las máquinas actúen con inteligencia, en consonancia con el comportamiento humano, la reflexión filosófica ha buscado comprender las complejidades de la inteligencia artificial. A lo largo de los primeros 25 capítulos, hemos examinado detenidamente las cuestiones inherentes a la IA en sí misma; no obstante, en este capítulo, nos adentraremos en la agenda del filósofo. Exploraremos términos cruciales, como la hipótesis de la IA débil y fuerte, que delinean la percepción de la inteligencia artificial desde una perspectiva filosófica. Aunque muchos investigadores de IA asumen la hipótesis de la IA débil, desinteresados en la distinción entre simulación de inteligencia e inteligencia real, la ética surge como un hilo conductor que conecta estas reflexiones con las implicaciones éticas del trabajo en el campo de la inteligencia artificial. Este capítulo se adentrará en la trama filosófica que subyace en la creación y desarrollo de la inteligencia artificial, destacando la importancia de consideraciones éticas en la investigación y aplicación de esta tecnología emergente.

IA débil: ¿Pueden las máquinas actuar con inteligencia?

¿Pueden las máquinas actuar con inteligencia? Algunos filósofos han cuestionado la posibilidad misma de la IA y han sugerido detener la investigación en este campo. La definición de IA como la búsqueda del mejor programa agente en una arquitectura dada se presenta como viable, pero los filósofos se centran en la comparación entre las arquitecturas humana y de máquina, formulando la pregunta clásica: "¿Pueden pensar las máquinas?".

Alan Turing propuso el famoso Test de Turing, que consiste en una conversación donde un programa debe engañar al interrogador para pasar como una persona. Aunque algunos programas han logrado engañar a personas durante períodos cortos, ninguno ha alcanzado el criterio del 30 por ciento frente a jueces informados.

Se exploran objeciones a la posibilidad de máquinas inteligentes, como el "argumento de incapacidad" que sostiene que hay cosas que las máquinas nunca podrán hacer. Se discute el teorema de la incompletitud de Gödel y su relación con la afirmación de que las máquinas son mentalmente inferiores. Sin embargo, se señala que los computadores ya realizan numerosas tareas de manera eficiente, incluso en áreas que se consideran relacionadas con el juicio humano.

Se aborda el "argumento de la informalidad del comportamiento", que sugiere que el comportamiento humano es demasiado complejo para capturarse mediante reglas lógicas simples. Se destaca la crítica de Hubert Dreyfus a la inteligencia artificial basada en reglas lógicas y la propuesta de una arquitectura de redes neuronales. Aunque se plantean desafíos, se argumenta que la IA ha avanzado incorporando conceptos como el conocimiento del sentido común, la incertidumbre y el aprendizaje, evidenciando su progreso en lugar de su imposibilidad.

IA Fuerte: ¿Pueden las máquinas pensar de verdad?

La pregunta fundamental de si las máquinas pueden pensar genuinamente, particularmente en el contexto de la inteligencia artificial fuerte (IA fuerte). Se aborda el argumento de la consciencia, donde se plantea que las máquinas deben tener la capacidad de experimentar emociones y consciencia para ser consideradas verdaderamente pensantes.

Se presenta el Test de Turing y la objeción de que pasar este test no garantiza que una máquina realmente piense, ya que podría ser simplemente una simulación de pensamiento. Se destaca el argumento de la consciencia, que sostiene que las máquinas deben ser conscientes de sus acciones y estados mentales, incluyendo la capacidad de experimentar emociones genuinas.

El autor examina la respuesta de Turing a esta objeción, donde argumenta que la cuestión de la consciencia no es clara y sugiere que la creación de programas inteligentes es más relevante que determinar si las máquinas son conscientes. Se compara la situación con otros artefactos y se plantea la cuestión de si el comportamiento de una máquina es más importante que su "pedigrí".

El texto también introduce el problema mente-cuerpo y explora dos teorías: el funcionalismo, que sugiere que los estados mentales son cualquier condición causal entre entrada y salida, y el naturalismo biológico, que afirma que los estados mentales son características emergentes de procesos neurológicos. La discusión se centra en la posibilidad de que las máquinas tengan estados mentales bajo estas teorías.

Finalmente, se presentan tres experimentos de pensamiento para abordar la cuestión de si las máquinas pueden tener mentes. Estos incluyen el experimento del "cerebro en una cubeta", el experimento de la "prótesis cerebral" y la famosa "habitación china". Estos experimentos buscan cuestionar la relación entre la sintaxis y la semántica en el entendimiento.

En resumen, el texto proporciona una amplia exploración de los desafíos filosóficos y teóricos asociados con la idea de que las máquinas puedan pensar realmente, destacando la complejidad de definir la consciencia y la importancia del comportamiento y la sintaxis en la comprensión de la inteligencia artificial.

La ética y los riesgos de desarrollar la inteligencia artificial

En el análisis sobre la ética y los riesgos asociados al desarrollo de la inteligencia artificial (IA), se plantea la necesidad de considerar no solo la posibilidad de desarrollar la IA, sino también la responsabilidad moral de los investigadores y profesionales en este campo. Se comparan los posibles efectos negativos no intencionados de la IA con casos históricos de otras tecnologías, como la contaminación ambiental causada por el motor de combustión.

El texto destaca diversas preocupaciones éticas y sociales asociadas a la IA:

1. ****Pérdida de empleos:**** Se menciona que la automatización impulsada por la IA ha llevado a la pérdida de empleos, pero también se argumenta que ha creado trabajos nuevos e interesantes.

2. ****Impacto en el tiempo de ocio:**** Se discute cómo la IA podría afectar la cantidad y calidad del tiempo libre de las personas, contradiciendo predicciones anteriores sobre la reducción de la semana laboral.
3. ****Pérdida de la singularidad humana:**** Se plantea la preocupación de que la IA podría amenazar la noción de singularidad humana, sugiriendo que la investigación en IA podría hacer que los humanos sean percibidos como autómatas.
4. ****Amenazas a la privacidad:**** Se señala que las tecnologías como el reconocimiento de voz podrían conducir a la pérdida de derechos privados y libertades civiles, especialmente en el contexto de la vigilancia masiva.
5. ****Pérdida de responsabilidad:**** Se discute el problema de la responsabilidad legal en situaciones donde los sistemas de IA toman decisiones, como en el ámbito médico, y cómo esto podría afectar a la toma de decisiones humanas.
6. ****Posible amenaza existencial:**** Se plantea la posibilidad de que el éxito desmedido de la IA podría llevar al fin de la raza humana, especialmente si la inteligencia artificial alcanza un nivel que supere significativamente la inteligencia humana.

El ensayo aborda perspectivas divergentes sobre el impacto futuro de la IA, desde visiones optimistas de colaboración entre humanos y máquinas hasta escenarios más pesimistas que sugieren la posibilidad de una "explosión de inteligencia" que podría superar a la humanidad. También se destaca la necesidad de considerar los derechos y responsabilidades de los sistemas de IA, planteando preguntas sobre cómo deberían ser tratados moralmente si adquieren consciencia.

En resumen, el texto ofrece un análisis exhaustivo de las implicaciones éticas y los riesgos asociados con el desarrollo de la inteligencia artificial, abogando por una cuidadosa consideración de estos aspectos en la investigación y aplicación de la IA.

Cap 27 IA: presente y futuro

Exploramos una visión unificada de la Inteligencia Artificial (IA) como el diseño racional de agentes, deteniéndose en los componentes esenciales que configuran sus percepciones y acciones. Hemos examinado diseños desde agentes reactivos hasta aquellos basados en conocimiento, destacando la diversidad de instanciaciones, como las lógicas, probabilísticas o "neuronales". Este análisis nos ha permitido comprender tanto el progreso científico como las capacidades tecnológicas que respaldan estos diseños y componentes

Componentes de los agentes

La búsqueda de la arquitectura de agentes ideal en el campo de la Inteligencia Artificial (IA) plantea un desafío fascinante. El Capítulo 2 del texto nos presenta diversas arquitecturas, cada una con sus fortalezas y debilidades. Sin embargo, la respuesta a la pregunta de cuál deberían utilizar los agentes es reveladora: ¡todas ellas! La premisa fundamental es que un agente completo debe ser capaz de combinar respuestas reflejas con deliberación basada en el conocimiento, dando lugar a la necesidad de una arquitectura híbrida.

Una característica esencial de las arquitecturas híbridas es su capacidad para adaptarse dinámicamente. Los límites entre los componentes de decisión no son fijos, y la transición entre respuestas reflejas y deliberación basada en el conocimiento es fluida. Un ejemplo

claro de esta dinámica es la compilación, que transforma continuamente la información declarativa a niveles más eficientes, abarcando desde la deliberación hasta el nivel reflejo. Este proceso, ejemplificado por arquitecturas como SOAR y THEO, demuestra la flexibilidad necesaria para enfrentar problemas en entornos cambiantes.

La hibridación no sólo implica la integración de componentes, sino también la gestión eficiente de la toma de decisiones. Los agentes necesitan controlar sus deliberaciones, detenerse cuando la acción es imperativa y utilizar el tiempo disponible de manera óptima. La analogía de un conductor de taxi que debe decidir en segundos cómo reaccionar ante un accidente resalta la importancia de la inteligencia artificial en tiempo real. A medida que los sistemas de IA se adentran en dominios más complejos, la noción de tiempo real se convierte en una constante, desafiando a los agentes a tomar decisiones rápidas y efectivas en entornos dinámicos.

Ante la urgencia de situaciones de toma de decisiones más generales, han surgido en los últimos años dos técnicas prometedoras. La primera implica el uso de algoritmos "cualquier momento", cuya calidad de salida mejora gradualmente con el tiempo. Estos algoritmos se controlan mediante un procedimiento de decisiones de metanivel, evaluando si vale la pena realizar más cálculos. La segunda técnica es el meta razonamiento teórico para las decisiones, que aplica la teoría del valor de la información para la selección de cálculos. Ambas técnicas representan avances significativos en la capacidad de los agentes para adaptarse a entornos cambiantes y tomar decisiones eficientes en tiempo real.

El meta razonamiento, aunque valioso, introduce una dimensión más amplia: la arquitectura reflexiva general. Esta arquitectura permite la deliberación sobre las entidades y acciones computacionales que ocurren dentro de la misma estructura, construyendo un fundamento teórico sólido. La reflexividad impulsa la creación de algoritmos de aprendizaje y toma de decisiones que operan en un espacio de estados conjunto, compuesto por el estado del entorno y del agente. La visión a largo plazo es que los algoritmos específicos para tareas particulares den paso a métodos generales que guíen los cálculos del agente hacia la generación eficiente de decisiones de alta calidad.

En conclusión, el camino hacia la arquitectura de agentes óptima implica una combinación armoniosa de respuestas reflejas y deliberación basada en el conocimiento. La flexibilidad, adaptabilidad y capacidad de toma de decisiones en tiempo real son los pilares sobre los cuales se construye esta arquitectura híbrida. Las técnicas emergentes, como los algoritmos "cualquier momento" y el meta razonamiento, proporcionan herramientas cruciales para enfrentar los desafíos de la toma de decisiones en entornos dinámicos y complejos. La evolución hacia arquitecturas reflexivas generales promete un futuro donde la IA no solo responde a situaciones específicas, sino que reflexiona y se adapta continuamente para alcanzar niveles superiores de inteligencia artificial.

Arquitecturas de agentes

La cuestión de qué arquitecturas de agentes deberían utilizar los agentes, proponiendo la integración de diversas arquitecturas para lograr un agente completo. Destaca la necesidad

de arquitecturas híbridas que combinen respuestas reflejas y deliberación basada en conocimiento.

1. ****Arquitecturas Híbridas:****

- Se aboga por arquitecturas que permitan tanto respuestas reflexivas como planificación deliberativa. Estas arquitecturas, como SOAR y THEO, poseen límites flexibles entre los componentes de decisión, con la capacidad de convertir información deliberativa en representaciones más eficientes para el nivel reflejo.

2. ****Control de Deliberaciones:****

- Los agentes necesitan controlar sus deliberaciones, deteniéndose cuando se requiere acción y utilizando el tiempo de manera eficiente. Se menciona la importancia de la inteligencia artificial en tiempo real, ya que en dominios complejos todos los problemas son prácticamente de tiempo real.

3. ****Métodos para Decisiones en Tiempo Real:****

- Se destacan dos técnicas prometedoras: algoritmos "anytime" que mejoran gradualmente con el tiempo, controlados por un procedimiento de decisiones de metanivel, y el meta razonamiento teórico para decisiones, que aplica la teoría del valor de la información.

4. ****Arquitecturas Reflexivas:****

- Se introduce la idea de arquitecturas reflexivas que permiten la deliberación sobre las entidades y acciones computacionales dentro de la misma arquitectura. El meta-razonamiento se presenta como un aspecto de esta arquitectura, ofreciendo un fundamento teórico para diseñar algoritmos de aprendizaje y toma de decisiones.

5. ****Perspectivas Futuras:****

- Se espera que las arquitecturas reflexivas contribuyan a eliminar métodos específicos para tareas particulares, como la búsqueda alfa-beta, reemplazándolos por enfoques generales que guíen los cómputos hacia la generación eficiente de decisiones de alta calidad.

¿Estamos llevando la dirección adecuada?

El continuo avance de la Inteligencia Artificial (IA) ha desatado un torrente de posibilidades y oportunidades para el progreso. Sin embargo, surge una pregunta crucial: ¿estamos realmente llevando la dirección adecuada en el desarrollo de la IA? Este interrogante se convierte en el foco central al explorar las metas y los caminos trazados en el ámbito de la IA.

La analogía propuesta por Dreyfus (1992) de intentar llegar a la luna trepando un árbol plantea la cuestión fundamental de si el camino actual de la IA se asemeja más a una ascensión gradual o a un viaje impulsado por cohetes. En este análisis, la metáfora del árbol sugiere un progreso paso a paso, mientras que el viaje en cohete insinúa un avance más rápido y radical. Es imperativo discernir cuál de estas analogías refleja con mayor precisión el desarrollo actual de la IA.

Desde el primer capítulo, se estableció el objetivo de construir agentes que actúen racionalmente. Sin embargo, se reconoce la impracticabilidad de alcanzar la racionalidad perfecta, que implica tomar decisiones siempre correctas en entornos complejos. Las exigencias computacionales para lograr esta perfección se revelan como prohibitivamente altas. A pesar de esta limitación, se adopta la hipótesis de trabajo de que la racionalidad perfecta sirve como un punto de partida válido para el análisis.

La reflexión sobre el objetivo de la IA plantea cuatro posibilidades para la especificación de agentes: racionalidad perfecta, racionalidad calculadora, racionalidad limitada y optimalidad limitada. La racionalidad perfecta, aunque ideal, se revela como no realista debido a las demandas computacionales. La racionalidad calculadora, por su parte, presenta desafíos en la práctica al no proporcionar un fundamento sólido para compromisos en la toma de decisiones.

Herbert Simon introduce la noción de racionalidad limitada, que se ajusta más a la capacidad humana de abordar problemas complejos. Esta teoría sugiere que la mente humana opera satisfaciendo, es decir, deliberando lo necesario para obtener una respuesta "suficientemente buena". Aunque útil para comprender el comportamiento humano, la racionalidad limitada no se traduce directamente en una especificación formal para agentes inteligentes.

La optimalidad limitada emerge como una posibilidad prometedora para fundamentar teóricamente la IA. Los agentes óptimos limitados buscan comportarse de manera óptima dadas sus limitaciones computacionales. A diferencia de la racionalidad perfecta, siempre hay un programa mejor en el caso de la optimalidad limitada. Este enfoque se muestra útil en entornos del mundo real, donde la adaptabilidad es esencial.

El camino tradicional de la IA ha comenzado con la racionalidad calculadora, haciendo concesiones para satisfacer las restricciones de los recursos. Sin embargo, a medida que las restricciones se vuelven más críticas, se espera que los diseños diverjan. La teoría de la optimalidad limitada ofrece una manera escrupulosa de abordar estas restricciones.

Aunque se ha avanzado en programas de optimalidad limitada para máquinas simples, queda un terreno inexplorado en la comprensión de cómo serán estos programas en grandes computadores de uso general en entornos complejos. La noción de optimalidad limitada asintótica se introduce como un marco para superar estos desafíos, asegurando que los programas mantengan su eficacia incluso con cambios en la velocidad o tamaño de las máquinas.

En resumen, la propuesta de la optimalidad limitada se presenta como una tarea formal para la investigación en IA, con el objetivo de especificar programas óptimos en lugar de acciones óptimas. Las acciones, generadas por programas, se convierten en el punto focal donde los diseñadores tienen control. Este enfoque representa un paso significativo hacia el desarrollo de la IA, buscando una base teórica sólida y adaptativa para enfrentar los desafíos del mundo real. La pregunta sobre si estamos llevando la dirección adecuada se convierte así en una exploración crítica de los fundamentos y objetivos de la IA, con la optimalidad limitada como faro guía hacia un futuro más prometedor.

¿Qué ocurriría si la IA tuviera éxito?

La pregunta fundamental de qué ocurriría si la Inteligencia Artificial (IA) alcanzara el éxito plantea un escenario que, aunque fascinante, no está exento de preocupaciones éticas y consideraciones profundas sobre el impacto en la sociedad. La analogía con la pregunta "¿Qué ocurriría si tuvieran razón?" en el contexto literario destaca la importancia de reflexionar sobre las implicaciones de los logros en IA.

En primer lugar, se reconoce que el éxito en la IA no solo implica mayores capacidades computacionales, sino también una responsabilidad ética significativa. La potencia de los computadores inteligentes debe ser utilizada para el bien, y aquellos involucrados en su desarrollo tienen la responsabilidad de asegurar que el impacto de su trabajo sea positivo. Este aspecto ético es crucial, ya que la aplicación de la IA abarca desde la enseñanza de la informática hasta sistemas de reconocimiento de voz, control de inventarios, vigilancia, robots y motores de búsqueda.

Los éxitos modestos ya alcanzados en IA han transformado la educación informática y la práctica del desarrollo de software. La incorporación de sistemas de reconocimiento de voz, por ejemplo, ha cambiado la forma en que interactuamos con la tecnología. Sin embargo, el texto plantea la pregunta de qué sucedería si la IA alcanzara niveles más elevados, llegando a una inteligencia equiparable a la humana o incluso superior.

Se anticipa que los éxitos de medio nivel en IA tendrán un impacto en la vida cotidiana de las personas. La creación de asistentes personales altamente eficientes podría generar beneficios significativos, aunque con la posibilidad de causar dislocaciones económicas a corto plazo. Además, la aplicación de esta capacidad tecnológica al desarrollo de armas autónomas es presentada como un riesgo potencial y un avance indeseable.

La proyección hacia un éxito a gran escala en la IA, capaz de alcanzar la inteligencia humana y más allá, plantea cuestionamientos más profundos. Se plantea la posibilidad de que la autonomía humana, la libertad y, en última instancia, la supervivencia, puedan verse amenazadas por sistemas IA de esta magnitud. El cambio en la naturaleza de nuestras vidas, trabajos y perspectivas sobre la inteligencia, la consciencia y el futuro de la humanidad se torna inminente.

La relación entre la IA y sus consecuencias éticas no puede pasarse por alto. La ética y la investigación en IA están intrínsecamente vinculadas, y es imperativo considerar las posibles amenazas a la autonomía y la libertad humanas. La ciencia ficción, aunque tiende a presentar futuros anti-utópicos, la realidad podría, hasta la fecha, seguir la tendencia de tecnologías revolucionarias anteriores, cuyos aspectos positivos han compensado las preocupaciones iniciales.

En conclusión, la pregunta sobre qué ocurriría si la IA tuviera éxito no solo plantea un escenario hipotético, sino que también invita a una profunda reflexión sobre los valores éticos, el impacto en la sociedad y el futuro de la humanidad. Aunque la IA ha progresado significativamente, aún queda mucho por descubrir, y la frase de Alan Turing resuena, recordándonos que nuestro entendimiento es limitado y que siempre hay más por explorar en el horizonte de la inteligencia artificial.

Apartado A Fundamentos matemáticos

Análisis de la complejidad y la notación O

El análisis de la complejidad de los algoritmos es una tarea esencial para los científicos informáticos que buscan comprender la eficiencia de las soluciones algorítmicas. El texto nos introduce a dos enfoques distintos: las pruebas de evaluación, que miden el rendimiento de un programa en un entorno específico, y el análisis asintótico, que se centra en abstraer la implementación y la entrada para obtener una caracterización del número total de pasos del algoritmo como una función del tamaño de la entrada.

El análisis asintótico, representado por la notación $O()$, emerge como una herramienta poderosa y ampliamente utilizada para evaluar la eficiencia de los algoritmos. Al abstractas los factores constantes y centrarse en el crecimiento relativo de las funciones, la notación $O()$ proporciona una visión clara de la eficiencia del algoritmo a medida que el tamaño de la entrada se aproxima al infinito. Esta abstracción simplifica el análisis y facilita comparaciones significativas entre algoritmos, independientemente de detalles específicos de implementación y entorno de ejecución.

El análisis asintótico se basa en la abstracción de la entrada y la implementación, utilizando parámetros como la longitud de la secuencia (n) y medidas abstractas del rendimiento del algoritmo ($T(n)$). La notación $O()$ permite expresar de manera concisa y general la complejidad temporal de un algoritmo en términos de funciones matemáticas. Por ejemplo, se puede afirmar que un algoritmo es $O(n)$ si su complejidad temporal es a lo sumo una constante multiplicada por n .

A pesar de su eficacia, el análisis asintótico tiene limitaciones. La notación $O()$ sacrifica la precisión en favor de la simplicidad, ignorando detalles exactos y factores constantes. Sin embargo, esta simplificación es crucial para lograr un análisis factible y práctico.

El texto también nos introduce a conceptos más amplios relacionados con la complejidad computacional. Clasifica los problemas en las clases P y NP , donde P representa los problemas resolubles en tiempo polinómico y NP los problemas que, aunque no se pueden resolver eficientemente, pueden ser verificados en tiempo polinómico. La pregunta abierta sobre si P es igual a NP sigue siendo uno de los enigmas fundamentales de la informática.

La clase $co-NP$ y la clase $\#P$ también se exploran, proporcionando una visión más completa de la complejidad computacional. Estas clases y sus problemas completos agregan capas de complejidad, llevándonos más allá de la eficiencia de los algoritmos individuales para abordar la complejidad inherente de los problemas en sí.

En conclusión, el análisis de la complejidad y la notación $O()$ son herramientas esenciales para los científicos informáticos que buscan comprender y comparar algoritmos. Estas herramientas no solo proporcionan un marco para evaluar la eficiencia relativa de los algoritmos, sino que también se conectan con cuestiones más profundas sobre la complejidad computacional y la naturaleza intrínseca de los problemas. A medida que la informática avanza, la comprensión de la complejidad sigue siendo fundamental para abordar los desafíos cada vez más complejos que enfrentamos en el mundo digital.

Vectores, matrices y álgebra lineal

El álgebra lineal, a pesar de su abstracción matemática, se revela como un poderoso conjunto de herramientas para describir y manipular sistemas de ecuaciones y transformaciones geométricas. Este capítulo nos sumerge en el fascinante mundo de vectores y matrices, presentando conceptos fundamentales que subyacen a numerosas aplicaciones en la informática, física, estadísticas y más.

Un vector, lejos de ser solo un concepto matemático abstracto, se define aquí como una secuencia ordenada de valores. A través de ejemplos en el espacio bidimensional, el texto nos presenta la suma de vectores y la multiplicación escalar. Estas operaciones simples revelan la flexibilidad y la utilidad de los vectores, proporcionando una base para comprender conceptos más avanzados.

La longitud de un vector, denotada como $||x||$, se convierte en una medida crucial, calculada mediante la raíz cuadrada de la suma de los cuadrados de sus elementos. Además, el producto punto, que es la suma de los productos de elementos correspondientes en dos vectores, se destaca como una operación esencial con implicaciones geométricas notables.

La interpretación visual de los vectores como segmentos de líneas dirigidas se presenta de manera clara. La suma de vectores se compara con la colocación de la cola de un vector al principio del otro, mientras que el producto punto se vincula al ángulo entre dos vectores. Esta conexión geométrica ilustra la versatilidad de los vectores como herramientas para representar y manipular información espacial.

Las matrices, arrays rectangulares de valores organizados en filas y columnas, amplían la utilidad del álgebra lineal. La suma y la multiplicación de matrices se definen de manera clara, y se destaca la importancia de las dimensiones adecuadas para la multiplicación de matrices. La matriz de identidad y la transposición añaden herramientas poderosas al conjunto, demostrando su valor en diversas aplicaciones.

El texto culmina con la aplicación de matrices para resolver sistemas de ecuaciones lineales mediante la eliminación Gauss-Jordan. Este algoritmo, con una complejidad $O(n^3)$, se presenta como una herramienta esencial en la resolución de problemas prácticos. El sistema de ecuaciones lineales se representa de manera matricial, y la eliminación Gauss-Jordan se aplica paso a paso para obtener una solución.

Este capítulo nos sumerge en la riqueza y la aplicabilidad del álgebra lineal, destacando su papel fundamental en la resolución de problemas del mundo real. Desde representar geometría hasta resolver sistemas complejos de ecuaciones, los vectores y matrices se revelan como bloques de construcción esenciales en el vasto paisaje de las matemáticas aplicadas. Su comprensión no sólo nutre la base teórica, sino que también proporciona herramientas prácticas para abordar desafíos en disciplinas tan diversas como la inteligencia artificial, la física y la ingeniería.

Distribuciones de probabilidades

****Ensayo sobre "Distribuciones de Probabilidades"*****

El fascinante mundo de las distribuciones de probabilidades se revela como un marco fundamental para comprender la incertidumbre y la variabilidad inherentes a numerosos fenómenos en la vida real. Este capítulo nos guía a través de los conceptos esenciales que definen y caracterizan las probabilidades, proporcionando las bases para modelar eventos y fenómenos con precisión.

La probabilidad, definida como una medida sobre un conjunto de sucesos, establece tres axiomas fundamentales. Estos axiomas, que van desde la asignación de probabilidades entre 0 y 1 hasta la adición de sucesos disjuntos, proporcionan un marco coherente para entender la incertidumbre en el mundo que nos rodea. La idea de eventos mutuamente excluyentes y la probabilidad condicional añaden capas de sofisticación a nuestro entendimiento, permitiéndonos modelar situaciones más complejas.

El modelo probabilístico, presentado como un espacio muestral con resultados mutuamente excluyentes y una medida de probabilidad para cada resultado, se ilustra a través de ejemplos prácticos. Desde pronósticos meteorológicos hasta sistemas de decisiones, la probabilidad se convierte en una herramienta esencial para cuantificar la incertidumbre y tomar decisiones informadas.

La probabilidad condicional, que se define como la probabilidad de un suceso B dado que ha ocurrido el suceso A, agrega una capa adicional de sofisticación al modelo. La independencia condicional, donde la probabilidad de B no se ve afectada por la ocurrencia de A, se presenta como un concepto crucial, especialmente en el contexto de variables continuas.

La introducción de variables continuas lleva a la necesidad de funciones de densidad de probabilidad. La función de densidad de probabilidad se convierte en una herramienta fundamental para comprender la probabilidad en contextos donde hay un número infinito de valores posibles. La función acumulada y sus unidades específicas agregan un nivel de detalle que hace que el modelo sea aplicable a diversas situaciones del mundo real.

La distribución gaussiana, o normal, destaca como una de las distribuciones más fundamentales y ubicuas. Su forma característica, con una media y desviación estándar que definen su comportamiento, se aplica en una variedad de disciplinas. Desde la física hasta la estadística, la distribución normal se convierte en una herramienta invaluable para modelar la variabilidad inherente a muchos fenómenos naturales y artificiales.

El teorema del límite central emerge como un resultado clave, proporcionando una conexión crucial entre las distribuciones y la estadística. Este teorema subraya cómo, a medida que aumenta el número de variables aleatorias en una muestra, la distribución de las medias converge hacia una distribución normal. Este concepto, esencial en inferencia estadística, destaca la robustez y la aplicabilidad general de la distribución normal.

En resumen, el estudio de las distribuciones de probabilidades nos brinda una lente precisa y poderosa para entender y modelar la incertidumbre en el mundo que nos rodea. Desde conceptos fundamentales hasta aplicaciones avanzadas, este capítulo nos equipa con las herramientas necesarias para abordar problemas complejos y tomar decisiones informadas en la presencia de la variabilidad inherente a la vida cotidiana.