IA: presente y futuro

El presente y futuro de la IA esta enfocándose en el diseño racional de agentes y su evolución. Diferentes tipos de agentes y sus posibles instanciaciones. Se ha logrado un avance significativo en el entendimiento científico y habilidades tecnológicas.

27.1 Componentes de los agentes

El párrafo 27.1 examina los componentes de agentes basados en utilidad y sus capacidades de aprendizaje. Se analizan avances y desafíos en áreas como interacción con el entorno, seguimiento del estado del mundo, proyección y selección de acciones futuras, utilidad como expresión de preferencias y aprendizaje. Aunque se han logrado avances en tecnología de sensores y actuadores, algoritmos de filtrado y percepción, y aprendizaje inductivo, aún existen desafíos en la construcción de representaciones de conocimiento a niveles de abstracción superiores y en la generación de funciones de utilidad realistas y preferencias complejas.

27.2 Arquitecturas de agentes

El párrafo 27.2 aborda las arquitecturas de agentes y cómo los agentes deberían utilizar todas ellas, incluyendo arquitecturas híbridas, para enfrentarse a diferentes situaciones. Los agentes necesitan controlar sus deliberaciones y adaptarse a entornos cambiantes. La inteligencia artificial en tiempo real y algoritmos de "cualquier momento" son técnicas prometedoras para abordar problemas en dominios más complejos. El metarazonamiento teórico para decisiones y arquitecturas reflexivas también pueden mejorar la calidad de las decisiones y la eficiencia computacional. En última instancia, se espera que los métodos específicos de tareas sean reemplazados por enfoques generales que generen decisiones de alta calidad de manera eficiente.

27.3 ¿Estamos llevando la dirección adecuada?

El párrafo discute el objetivo de la Inteligencia Artificial (IA) y plantea cuatro posibilidades para abordar la especificación de agentes: Racionalidad perfecta, Racionalidad calculadora, Racionalidad limitada y Optimalidad limitada. La racionalidad perfecta no es realista debido a las demandas computacionales, y la racionalidad calculadora no proporciona una base teórica sólida para hacer compromisos. La racionalidad limitada, propuesta por Herbert Simon, es un modelo útil para el comportamiento humano, pero carece de una especificación formal. La optimalidad limitada, que tiene en cuenta los recursos computacionales, parece ser la mejor opción para un fundamento teórico sólido en IA. Aunque se han hecho avances en el estudio de la optimalidad limitada, aún queda mucho por descubrir. Se sugiere que una ciencia de IA basada en la optimalidad limitada se centraría en los procesos que permiten a un programa de agentes converger hacia la optimalidad, en lugar de centrarse en los detalles de los programas complicados resultantes.

27.4 ¿Qué ocurriría si la IA tuviera éxito?

El párrafo reflexiona sobre las posibles consecuencias del éxito en la Inteligencia Artificial (IA) y las responsabilidades éticas de los investigadores en este campo. Los avances en IA ya han permitido aplicaciones como reconocimiento de voz, control de inventarios, vigilancia, robots y motores de búsqueda. Se espera que éxitos de mayor nivel afecten la vida cotidiana de las personas. Aunque la IA tiene el potencial de impactar positivamente la vida humana, también presenta preocupaciones, como la creación de armas autónomas y la amenaza a la autonomía, libertad y supervivencia humanas. La investigación en IA no puede separarse de sus consecuencias éticas y se debe tener en cuenta tanto los aspectos positivos como los negativos de la tecnología. Actualmente, la IA ha avanzado en su corta existencia, pero aún queda mucho por hacer, como señaló Alan Turing en su ensayo "Computing Machinery and Intelligence".