Entrevista a Ramón López de Mántaras

Ricardo Morte Ferrer
(LI²FE)

Aníbal Monasterio Astobiza
(Oxford University/LI²FE)

Introducción: Ramón López de Mántaras i Badia (Sant Vicenç de Castellet, 1952) es informático y físico. Doctor en Física por la Universidad Paul Sabatier de Toulouse (Francia) Master of Science en Informática por la Universidad de California Berkeley y Doctor en Informática por la Universitat Politècnica de Catalunya. Investigador del Consejo Superior de Investigaciones Científicas (CSIC), Director del Instituto de Investigación en Inteligencia Artificial y "Chancellor's Distinguished Visiting Professor" de la "University of Technology Sydney". Profesor externo de la "University of Technology Sydney" y de la "Western Sydney University". Es uno de los pioneros de la inteligencia artificial en España. En el año 2011 recibió el Premio Robert S. Engelmore de la Asociación Americana de Inteligencia Artificial, en el 2012 el "Premio Nacional de Informática" de la "Sociedad Científica Informática" de España y en el 2016 el EurAI Distinguished Service Award de la "European Association of Artificial Intelligence". Actualmente sus principales campos de investigación son el razonamiento basado en casos, los robots autónomos capaces de aprender con su entorno físico ("Developmental Robotics") y la inteligencia artificial aplicada a la música.

ISSN 1989-7022

¿Cómo es el día a día del director del IIIA (Instituto de Investigación en Inteligencia Artificial) ubicado en Barcelona?

En mi caso, gracias a una gerente y una administración extraordinariamente eficientes, no es muy distinto que el de otros investigadores sénior. Obviamente cuando hay problemas, como por ejemplo los relacionados con la enorme rigidez y control administrativo que el MAP y el Ministerio de Hacienda imponen al CSIC, no puedo evitar sufrir estrés y sentir impotencia ante tanta insensatez. En lugar de dedicarnos a lo nuestro, es decir investigar, nos hacen perder cientos de horas cada año, por ejemplo, justificando repetidas veces los gastos derivados de los proyectos. Si la gestión administrativa fuera sensata, como la que se aplica en los países anglosajones, ser Director sería mucho más fácil e investigar también.

¿Qué hacéis en vuestro instituto y cuántas personas trabajan allí?

Investigamos en tres grandes temas de la Inteligencia Artificial: Sistemas Multiagente, Aprendizaje Automático, Razonamiento Automático, Resolución de problemas mediante modelos basados en Restricciones y Modelos Lógico-matemáticos para la IA. Actualmente somos alrededor de 60 personas, una tercera parte de las cuales somos permanentes.

ISSN 1989-7022
DILEMATA, año 9 (2017), nº 24, 301-309

¿Qué proyectos tienes en mente o en qué estás trabajando ahora?

Actualmente estoy trabajando en métodos para acelerar los algoritmos de Aprendizaje por Refuerzo - *Reinforcement Learning* (RL) en inglés -. Es un tipo de aprendizaje que es común en humanos y otros animales y es de gran utilidad en IA y en particular en robótica. Uno de los principales problemas computacionales de este aprendizaje es que es lento, debido a que el proceso de entrenamiento del sistema es demasiado lento cuando el conjunto de estados y acciones del sistema es grande. En colaboración con otros investigadores estamos incorporando heurísticas a estos algoritmos de RL para acelerarlos. Hemos obtenido resultados muy buenos aplicando heurísticas basadas en Razonamiento Basado en Casos que se basan en resolver problemas mediante el uso de la experiencia adquirida resolviendo problemas anteriores similares. También hemos aplicado estas ideas de usar experiencia previa resolviendo problemas a otra área muy importante en Aprendizaje que en inglés se llama "*Transfer Learning*" y consiste en aprovechar lo que se ha aprendido resolviendo una tarea dada en un dominio de aplicación para reusarlo a la hora de aprender a llevar a cabo más rápidamente otra tarea distinta más compleja en lugar de tener que volver a aprender desde cero. Por ejemplo, lo hemos aplicado en el campo de la Robótica Humanoide de la forma siguiente. Nuestro algoritmo primero aprendió a mantener en equilibrio un doble péndulo invertido y a continuación las acciones aprendidas sirvieron para que un robot humanoide aprendiera más rápidamente a mantenerse en equilibrio estable mientras caminaba.

¿Qué aspectos de la sociedad crees que se pueden mejorar con el uso de sistemas y técnicas de la IA?

En principio prácticamente en todos los aspectos, desde la salud hasta el comercio electrónico pasando por la movilidad, el ocio o la práctica de actividades artísticas. En estos y muchos otros campos existen aplicaciones muy impresionantes. Gracias a la disponibilidad de analizar enormes cantidades de datos mediante, por ejemplo, aprendizaje profundo, es posible diagnosticar con gran precisión (en ocasiones mejor que los médicos) e incluso pronosticar enfermedades con más antelación que los médicos. También hemos visto recientemente espectaculares resultados en juegos como Go o póker y no digamos en ajedrez! Estos sistemas son magníficos "*coaches*" para jugadores humanos. En movilidad, los futuros coches cada vez más autónomos harán bajar significativamente los accidentes de tráfico y cambiarán

drásticamente el concepto actual de movilidad basada principalmente en el coche individual. Dejaremos de ser propietarios de coches para ser usuarios de flotas de coches autónomos que prácticamente no pararán de transportar personas y mercancías de un lugar a otro. Actualmente nuestros coches pasan unas 24 horas al día aparcados, lo cual es bastante absurdo. Cuando no se requiera un conductor humano no tendrán porque estar aparcados casi todo el día. En ese momento el concepto de propiedad, "mi coche", desaparecerá. En la actividad artística cada vez es más necesario el uso de la informática como herramienta de ayuda a la creación, con la IA se pasará de que el ordenador sea una herramienta pasiva a que trabaje en equipo con artistas, diseñadores, arquitectos, músicos, etc. y el resultado de ese binomio será superior al de cada uno de los componentes (persona y máquina) por separado. Es decir que hay un fenómeno de sinergia. El motivo es que si consideramos el acto creativo como la combinación de elementos existentes de forma novedosa e interesante (por ejemplo, en música los elementos serían las notas musicales, su duración, su dinámica, su articulación, la harmonía, etc.) entonces un software basado en IA de ayuda a la composición puede explorar muy rápidamente muchas más combinaciones de estos elementos que un compositor humano y proponer al compositor aquellas combinaciones que pudieran ser más interesantes para un estilo musical dado. Lo mismo en artes visuales, diseño de nuevos productos, etc.

¿Nos podrías explicar en términos muy sencillos cuáles son las diferencias entre "aprendizaje máquina" y "aprendizaje profundo", qué son las "redes neuronales" y decirnos qué pueden aprender y hacer a través de estas técnicas a día de hoy las máquinas?

El aprendizaje máquina es un concepto más genérico que incluye, entre otras técnicas, a las técnicas de aprendizaje profundo. La técnica más extendida de aprendizaje profundo son las redes neuronales convolucionales con un gran número de capas intermedias. Hasta no hace mucho las redes neuronales se diseñaban con un número muy limitado de capas intermedias debido a la dificultad de entrenar dichas redes para que aprendan correctamente y eficientemente. Cuantas más capas tiene la red más datos de entrenamiento se requieren y más tardan en aprender, actualmente gracias a la disponibilidad de grandes cantidades de datos en muchos ámbitos de aplicación y gracias al acceso a sistemas de computación de altas prestaciones basados en clústeres de GPUs es posible entrenar correctamente y eficientemente

ISSN 1989-7022 DILEMATA, año 9 (2017), nº 24, 301-309

ISSN 1989-7022
DILEMATA, año 9 (2017), nº 24, 301-309

redes neuronales con un elevado número de capas. Lo que estas redes pueden aprender es a reconocer patrones y en particular patrones presentes en series temporales y en imágenes bidimensionales de ahí los grandes éxitos obtenidos en aplicaciones al diagnóstico médico basado en imágenes, al reconocimiento de objetos, o al reconocimiento del habla. Los datos que procesan los sistemas de aprendizaje profundo consisten en vectores y matrices de números y por ese motivo se trata de técnicas sub-simbólicas. El aprendizaje máquina incluye además técnicas de aprendizaje simbólico capaces de aprender en base a representaciones de datos consistentes en símbolos de más alto nivel como por ejemplo fórmulas lógicas.

¿Cuál es el futuro de la IA y qué aplicaciones ves para el futuro?

Durante mucho tiempo la IA seguirá siendo IA débil, es decir que serán IAs que sabrán resolver muy bien problemas muy concretos y específicos aunque avanzaremos lentamente hacia la consecución de IAs cada vez más generales y versátiles. Posiblemente la lección más importante que hemos aprendido a lo largo de los 60 años de existencia de la inteligencia artificial es que lo que parecía más difícil (diagnosticar enfermedades, jugar al ajedrez, Go o póker al más alto nivel, etc.) ha resultado ser relativamente fácil y lo que parecía más fácil ha resultado ser tan difícil que todavía no lo hemos conseguido. Las capacidades más complicadas de alcanzar son aquellas que requieren interaccionar con entornos no restringidos: percepción visual, comprensión del lenguaje, razonar con sentido común y tomar decisiones con información incompleta. Diseñar sistemas que tengan estas capacidades requiere integrar desarrollos en muchas áreas de la Inteligencia Artificial. En particular, necesitamos lenguajes de representación de conocimientos que codifiquen información acerca de muchos tipos distintos de objetos, situaciones, acciones, etc., así como de sus propiedades y de las relaciones entre ellos. También necesitamos nuevos algoritmos que, en base a estas representaciones, puedan responder de forma robusta y eficiente preguntas sobre prácticamente cualquier tema. Finalmente, dado que necesitarán conocer un número prácticamente ilimitado de cosas, estos sistemas deberán ser capaces de aprender nuevos conocimientos de forma continua a lo largo de toda su existencia. En definitiva, además de progresos individuales en cada una de estas áreas, el futuro de la IA pasa por diseñar sistemas que integren percepción, representación, razonamiento, acción y aprendizaje. Éste es un problema muy importante en IA ya que todavía no sabemos cómo integrar todos estos componentes de la inteligencia. Necesitamos arquitecturas cognitivas que integren a estos componentes de forma

adecuada. Los sistemas integrados son un paso previo fundamental para conseguir algún día inteligencias artificiales de tipo general.

Entre las actividades futuras, creo que los temas de investigación más importantes seguirán siendo el aprendizaje automático, los sistemas multiagente, el razonamiento espacial, la planificación de acciones, el razonamiento basado en la experiencia, la visión artificial, la comunicación multimodal persona-máquina y la robótica humanoide y animaloide y en particular la robótica basada en el desarrollo mental ("*developmental robotics*"). En el caso de la robótica, veremos progresos significativos gracias a las aproximaciones biomiméticas para reproducir en máquinas el comportamiento de animales. De hecho no se trata únicamente de reproducir el comportamiento de un animal sino también de comprender como funciona el cerebro que produce dicho comportamiento. Se trata de construir y programar circuitos electrónicos que reproduzcan las secuencias de órdenes que el cerebro genera para producir los movimientos (de las alas, las patas, etc.). Los biólogos están interesados en los intentos de fabricar un cerebro artificial porque es una manera de comprender mejor el órgano y los ingenieros buscan información biológica para hacer diseños más eficaces. Mediante la biología molecular es posible identificar qué genes y que neuronas juegan un papel en estos movimientos.

En cuanto a las aplicaciones, las más importantes serán aquellas relacionadas con la web, los videojuegos, los robots autónomos. La economía y la sociología también usarán cada vez más modelos de IA, en particular modelos basados en agentes para simular interacciones entre grandes cantidades de agentes y predecir, por ejemplo, posibles situaciones de crisis. La creatividad artística se intensificará gracia a la IA. También los descubrimientos científicos se beneficiarán de la IA, por ejemplo en biología molecular y farmacología veremos un uso cada vez más importante de la IA. Muchos fármacos tienen efectos secundarios inesperados que son a menudo beneficiosos. En lugar de esperar que estos efectos secundarios se descubran por casualidad, los investigadores en farmacología ya aplican técnicas de IA para predecir que fármacos existentes pueden tener otros usos terapéuticos con la consiguiente ganancia de tiempo que ello supone frente al desarrollo y aprobación de un nuevo fármaco.

Hay muchas voces optimistas sobre los logros que puede conseguir la investigación en IA. Y al mismo tiempo que son optimistas nos alertan de los potenciales peligros de una Inteligencia Artificial descontrolada. Nos hablan

ISSN 1989-7022 DILEMATA, año 9 (2017), nº 24, 301-309

ISSN 1989-7022

DILEMATA, año 9 (2017), nº 24, 301-309

del desempleo de millones de personas a consecuencia de la automatización y la introducción de las máquinas en el trabajo e incluso de que se pueden rebelar contra sus creadores (nosotros los seres humanos). Hablan de la idea de Singularidad defendida vehemente por el inventor, científico, empresario y director de ingeniería en Alphabet Ray Kurzweil, de superinteligencia, concepto acuñado por el filósofo Nick Bostrom que alerta que algún día las máquinas pueden suponer una amenaza y riesgo existencial para la humanidad. ¿Qué dilemas éticos puede plantear que las máquinas lleguen a ser más inteligentes que los seres humanos y si son realistas estos vaticinios?

En mi opinión estos vaticinios son poco realistas pues suponen una predicción muy exagerada que no se sustenta en la realidad del estado del arte de la IA. A pesar de impresionantes éxitos recientes como por ejemplo el software AlphaGo que batió ampliamente a Lee Sedol, campeón mundial de Go, o Libratus que batió a expertos jugadores de póker en la modalidad "*no-limit Texas Hold'em*", actualmente todavía nos encontramos con importantes dificultades para que una máquina comprenda realmente frases relativamente sencillas o sepa interpretar el significado de lo que ve. La comprensión profunda del lenguaje y de las escenas que observamos solamente es posible si, entre otras cosas, poseemos conocimientos de sentido común. La adquisición de conocimientos de sentido común es el principal problema al que se enfrenta la Inteligencia Artificial. Poseer sentido común es el requerimiento fundamental para que las máquinas actuales hagan el salto cualitativo de tener inteligencias artificiales específicas (que les permiten saber hacer muy bien una única cosa muy bien definida y restringida) y empiecen a tener inteligencia artificial de tipo general, similar a la inteligencia humana. Los conocimientos de sentido común, necesarios para dar este salto cualitativo, no se suelen recoger en libros o enciclopedias y sin embargo todos los seres humanos poseemos una gran cantidad de ellos. Una aproximación a la adquisición de conocimiento de sentido común es la denominada "cognición situada". Es decir, situar a la máquina en entornos reales, como ocurre con los seres humanos, con el fin de que tengan "vivencias" y experiencias que les doten de sentido común mediante un mecanismo de aprendizaje basado en el desarrollo mental en el sentido de Piaget. Esta cognición situada requiere que la IA forme parte de un cuerpo con el que interaccionar con el entorno. En el caso de los humanos y otros animales, los cerebros forman parte integrante de sus cuerpos que, a su vez, están situados e interaccionan en un entorno real muy complejo. De hecho basamos una gran parte

de nuestra inteligencia en nuestra capacidad sensorial y motora. En otras palabras, el cuerpo da forma a la inteligencia ("*the body shapes the way we think*") y por lo tanto sin cuerpo no puede haber inteligencia general completa. Esto es así porque el "hardware" del cuerpo, en particular los detalles del sistema sensor y del sistema motor, determina el tipo de situaciones que un agente puede percibir y abordar. A su vez, estas situaciones conforman las habilidades cognitivas de los agentes. Consecuentemente, para especificar concretamente dichas habilidades cognitivas es necesario tener en cuenta las interacciones del agente con su entorno.

Poniendo ejemplos concretos, ¿qué les dirías a quienes afirman que la IA no es de fiar? Entre otros motivos mencionan que los programadores/ desarrolladores de algoritmos terminan por proyectar algunos de sus prejuicios en esos algoritmos.

Idealmente lo sistemas de IA que tomen decisiones, de forma completamente autónoma, que pueden afectar significativamente a la sociedad deberían tener "valores" alineados con los valores humanos y de hecho recientemente ha emergido una línea de investigación en este sentido pero el problema es ¿qué valores? Diseñadores distintos pueden tener valores distintos. Este problema debería hacernos reflexionar sobre la conveniencia de dotar de autonomía absoluta a las máquinas. En mi opinión en la gran mayoría de casos nunca deberíamos eliminar al ser humano del proceso de decisión.

¿No crees que un alto grado de transparencia es esencial en la implantación de sistemas de IA? En principio eso deberá permitir que se refuercen sus beneficios y se reduzcan sus riesgos.

Sin duda, aspectos como la transparencia, la rendición de cuentas, la responsabilidad, etc. deberían ser claves a la hora de implantar sistemas basados en IA. En la reciente "Declaración de Barcelona para un desarrollo y uso adecuados de la Inteligencia Artificial en Europa" (http//www.iiia.csic.es/barcelonadeclaration) hemos abordado estos y otros aspectos.

¿Crees que estamos cerca de alcanzar el sueño de Leibniz de convertir las leyes que gobiernan el razonamiento y la razón humana en un ejercicio computacional y por consiguiente ser capaces de crear mentes digitales?

ISSN 1989-7022

ISSN 1989-7022
DILEMATA, año 9 (2017), nº 24, 301-309

En mi opinión a corto e incluso medio plazo no creo. Antes he hablado del problema de sentido común. Creo que es precisamente el obstáculo principal para alcanzar el sueño de Leibniz. Ese es de hecho el objetivo de la IA "fuerte". Quien introdujo esta distinción entre IA débil y IA fuerte fue el filósofo John Searle en un artículo crítico con la IA publicado en 1980 que provocó, y sigue provocando, mucha polémica. La IA fuerte implicaría que un ordenador convenientemente programado no simula una mente sino que *es una mente* y por consiguiente debería ser capaz de pensar igual que un ser humano. La IA fuerte está relacionada con la hipótesis del Sistema de Símbolos Físicos formulada por Allen Newell y Herbert Simon (dos de los fundadores de la IA en 1956) en una ponencia con motivo de la recepción del prestigioso "Premio Turing" en 1975. Según esta hipótesis, todo sistema de símbolos físicos posee los medios necesarios y suficientes para llevar a cabo acciones inteligentes. Por otra parte dado que los seres humanos somos capaces de mostrar conductas inteligentes en el sentido general, entonces, de acuerdo con la hipótesis, nosotros somos también sistemas de símbolos físicos. Conviene aclarar a que se refieren Newell y Simon cuando hablan de Sistema de Símbolos Físicos (SSF). Un SSF consiste en un conjunto de entidades denominadas símbolos que, mediante relaciones, pueden ser combinados formando estructuras más grandes - como los átomos que se combinan formando moléculas – y que pueden ser transformados aplicando un conjunto de procesos. Estos procesos pueden crear nuevos símbolos, crear y modificar relaciones entre símbolos, almacenar símbolos, comparar si dos símbolos son iguales o distintos, etcétera. Estos símbolos son físicos en tanto que tienen un substrato físico-electrónico (en el caso de los ordenadores) o físico-biológico (en el caso de los seres humanos). Efectivamente, en el caso de los ordenadores los símbolos se realizan mediante circuitos electrónicos digitales y en el caso de los seres humanos mediante redes de neuronas. En definitiva, de acuerdo con la hipótesis SSF, la naturaleza del substrato (circuitos electrónicos o redes neuronales) carece de importancia siempre y cuando dicho substrato permita procesar símbolos. No olvidemos que se trata de una hipótesis y por lo tanto no debe de ser ni aceptada ni rechazada a priori. En cualquier caso su validez o refutación se deberá verificar, de acuerdo con el método científico, con ensayos experimentales. La IA es precisamente el campo científico dedicado a intentar verificar esta hipótesis en el contexto de los ordenadores digitales, es decir verificar si un ordenador convenientemente programado es capaz o no de tener conducta inteligente de tipo general.

En cualquier caso, por muy inteligentes que lleguen a ser las futuras inteligencias artificiales, de hecho siempre serán distintas a las inteligencias humanas ya que las inteligencias dependen de los cuerpos en los que están situadas. Eso es así debido a que el desarrollo mental que requiere toda inteligencia compleja depende de las interacciones con el entorno y estas interacciones dependen a su vez del cuerpo, en particular del sistema perceptivo y del sistema motor (recordemos lo dicho anteriormente de que "*the body shapes the way we think*"). El hecho de ser inteligencias ajenas a la humana y por lo tanto ajenas a los valores y necesidades humanas nos debería hacer reflexionar sobre posibles limitaciones éticas al desarrollo de la Inteligencia Artificial.

ISSN 1989-7022 ***DILEMATA**, año 9 (2017), nº 24, 301-309*