Inteligencia Artificial

Retos actuales y participación de Inria

Inria | LIBRO BLANCO | N°01 | SEGUNDA EDICIÓN - 2021 |

Del original:
Bertrand Braunschweig. Artificial Intelligence: Current challenges and Inria's engagement - Inria white paper. INRIA, Livre blanc Inria N°1, pp.154, 2021. hal-01564589v2
2º Edición
Fecha de publicación: 2021, Francia

Traducción autorizada del idioma inglés de la edición publicada por Inria

Traducción al español realizada por Inria Chile

Coordinación general y revisión técnica: Nayat Sánchez Pi

Coordinación de producción: Julia Allirot y Katherine Lippi

Coordinación ejecutiva: Andrés Vignaga

Revisión gráfica y maquetación: Estudio Paretti y Katherine Lippi

Agradecimientos – Gracias a las siguientes personas de Inria Chile por sus aportes a la revisión técnica: Hugo Carrillo, Hernán Lira, Luis Martí, Luis Valenzuela

Impreso por Lahosa
Impresión: Octubre 2022, Chile
ISBN: 978-956-09873-0-3

Impreso en Chile / Printed in Chile

Índice

0. Investigadores de los equipos de proyectos y centros de Inria que contribuyeron a la elaboración de este documento (fueron entrevistados, aportaron textos o ambas cosas)[1]

Abiteboul Serge*, integrante del antiguo equipo-proyecto DAHU, Saclay
Alexandre Frédéric**, jefe del equipo-proyecto MNEMOSYNE, Burdeos
Altman Eitan**, integrante del equipo-proyecto NEO, Sophia-Antipolis
Amsaleg Laurent**, jefe del equipo-proyecto LINKMEDIA, Rennes
Antoniu Gabriel**, jefe del equipo-proyecto KERDATA, Rennes
Arlot Sylvain**, jefe del equipo-proyecto CELESTE, Saclay
Ayache Nicholas***, jefe del equipo-proyecto EPIONE, Sophia-Antipolis
Bach Francis***, jefe del equipo-proyecto SIERRA, París
Beaudouin-Lafon Michel**, integrante del equipo-proyecto EX-SITU, Saclay
Beldiceanu Nicolas*, jefe del antiguo equipo-proyecto TASC, Nantes
Bellet Aurélien**, responsable de la acción exploratoria FLAMED, Lille
Bezerianos Anastasia **, integrante del equipo-proyecto ILDA, Saclay
Bouchez Florent**, responsable de la acción exploratoria AI4HI, Grenoble
Boujemaa Nozha*, ex asesor en materia de big data del Presidente de Inria
Bouveyron Charles**, jefe del equipo-proyecto MAASAI, Sophia-Antipolis
Braunschweig Bertrand***, director, coordinación del programa nacional de investigación sobre IA
Brémond François***, jefe del equipo-proyecto STARS, Sophia-Antipolis
Brodu Nicolas**, responsable de la acción exploratoria TRACME, Burdeos
Cazals Frédéric**, jefe del equipo-proyecto ABS, Sophia-Antipolis
Casiez Géry**, integrante del equipo-proyecto LOKI, Lille
Charpillet François***, jefe del equipo-proyecto LARSEN, Nancy
Chazal Frédéric**, jefe del equipo-proyecto DATASHAPE, Saclay y Sophia-Antipolis
Colliot Olivier***, jefe del equipo-proyecto ARAMIS, París
Cont Arshia*, jefe del antiguo equipo-proyecto MUTANT, París
Cordier Marie-Odile*, integrante del equipo-proyecto LACODAM, Rennes
Cotin Stephane**, jefe del equipo-proyecto MIMESIS, Estrasburgo
Crowley James***, ex jefe del equipo-proyecto PERVASIVE, Grenóble
Dameron Olivier**, jefe del equipo-proyecto DYLISS, Rennes
De Charette, Raoul**, integrante del equipo-proyecto RITS, París

[1] (*): primera edición, 2016; (**): segunda edición, 2020; (***) ambas ediciones

De La Clergerie Eric*, integrante del equipo-proyecto ALMANACH, París
De Vico Fallani Fabrizio*, integrante del equipo-proyecto ARAMIS, París
Deleforge Antoine**, responsable de ACOUST. Acción exploratoria IA2, Nancy
Derbel Bilel**, integrante del equipo-proyecto BONUS, Lille
Deriche Rachid**, jefe del equipo-proyecto ATHENA, Sophia-Antipolis
Dupoux Emmanuel**, jefe del equipo-proyecto COML, París
Euzenat Jérôme***, jefe del equipo-proyecto MOEX, Grenoble
Fekete Jean-Daniel**, jefe del equipo-proyecto AVIZ, Saclay
Forbes Florence**, jefe del equipo-proyecto STATIFY, Grenoble
Franck Emmanuel**, responsable de la acción exploratoria MALESI, Nancy
Fromont Elisa, **, responsable del desafío HYAIAI Inria, Rennes
Gandon Fabien***, jefe del equipo-proyecto WIMMICS, Sophia-Antipolis
Giavitto Jean-Louis*, integrante del antiguo equipo-proyecto MUTANT, París
Gilleron Rémi*, integrante del equipo-proyecto MAGNET, Lille
Giraudon Gérard*, ex director del centro de investigación Sophia-Antipolis Méditerranée
Girault Alain**, director científico adjunto
Gravier Guillaume*, ex jefe del equipo-proyecto LINKMEDIA, Rennes
Gribonval Rémi**, integrante del equipo-proyecto DANTE, Lyon
Gros Patrick*, director del centro de investigación Grenoble-Rhône Alpes
Guillemot Christine**, jefa del equipo-proyecto SCIROCCO, Rennes
Guitton Pascal*, integrante del equipo-proyecto POTIOC, Burdeos
Horaud Radu***, jefe del equipo-proyecto PERCEPTION, Grenoble
Jean-Marie Alain**, jefe del equipo-proyecto NEO, Sophia-Antipolis
Laptev Ivan**, integrante del equipo-proyecto WILLOW, París
Legrand Arnaud**, jefe del equipo-proyecto POLARIS, Grenoble
Lelarge Marc**, jefe del equipo-proyecto DYOGENE, París
Mackay Wendy**, jefe del equipo-proyecto EX-SITU, Saclay
Malacria Sylvain**, integrante del equipo-proyecto LOKI, Lille
Manolescu Ioana*, jefe del equipo-proyecto CEDAR, Saclay
Mé Ludovic**, director científico adjunto
Merlet Jean-Pierre**, jefe del equipo-proyecto HEPHAISTOS, Sophia-Antipolis
Maillard Odalric-Ambrym**, responsable de la acción exploratoria SR4SG, Lille
Mairal Julien**, jefe del equipo-proyecto THOTH, Grenoble
Moisan Sabine*, integrante del equipo-proyecto STARS, Sophia-Antipolis
Moulin-Frier Clément**, responsable de la acción exploratoria ORIGINS, equipo-proyecto FLOWERS, Burdeos
Mugnier Marie-Laure***, jefa del equipo-proyecto GRAPHIK, Montpellier
Nancel Mathieu**, integrante del equipo-proyecto LOKI, Lille
Nashashibi Fawzi***, jefe del equipo-proyecto RITS, París

Neglia Giovanni**, responsable de la acción exploratoria MAMMALS, Sophia-Antipolis
Niehren Joachim*, jefe del equipo-proyecto LINKS, Lille
Norcy Laura**, asociaciones europeas
Oudeyer Pierre-Yves***, jefe del equipo-proyecto FLOWERS, Burdeos
Pautrat Marie-Hélène**, directora de asociaciones europeas
Pesquet Jean-Christophe**, jefe del equipo-proyecto OPIS, Saclay
Pietquin Olivier*, ex integrante del equipo-proyecto SEQUEL, Lille
Pietriga Emmanuel**, jefe del equipo-proyecto ILDA, Saclay
Ponce Jean*, jefe del equipo-proyecto WILLOW, París
Potop Dumitru**, integrante del equipo-proyecto KAIROS, Sophia-Antipolis
Preux Philippe***, jefe del equipo-proyecto SEQUEL (SCOOL), Lille
Roussel Nicolas***, director del centro de investigación Bordeaux Sud Ouest
Sagot Benoit***, jefe del equipo-proyecto ALMANACH, París
Saut Olivier**, jefe del equipo-proyecto MONC, Burdeos
Schmid Cordelia*, ex jefa del equipo-proyecto THOTH, Grenoble, actualmente integrante del equipo-proyecto WILLOW, París
Schoenauer Marc***, codirector del equipo-proyecto TAU, Saclay
Sebag Michèle***, codirectora del equipo-proyecto TAU, Saclay
Seddah Djamé*, integrante del equipo-proyecto ALMANACH, París
Siegel Anne***, ex jefa del equipo-proyecto DYLISS, Rennes
Simonin Olivier***, jefe del equipo-proyecto CHROMA, Grenoble
Sturm Peter*, director científico adjunto
Termier Alexandre***, jefe del equipo-proyecto LACODAM, Rennes
Thiebaut Rodolphe**, jefe del equipo-proyecto SISTM, Burdeos
Thirion Bertrand**, jefe del equipo-proyecto PARIETAL, Saclay
Thonnat Monique*, integrante del equipo-proyecto STARS, Sophia-Antipolis
Tommasi Marc***, jefe del equipo-proyecto MAGNET, Lille
Toussaint Yannick*, integrante del equipo-proyecto ORPAILLEUR, Nancy
Valcarcel Orti Ana**, coordinación del programa nacional de investigación en IA
Vercouter Laurent**, coordinación del programa nacional de investigación en IA
Vincent Emmanuel***, integrante del equipo-proyecto MULTISPEECH, Nancy

1

Samuel y su Mayordomo

Samuel y su mayordomo[2]

A las 07:15 horas, Sam se despierta y se prepara para un día normal de trabajo. Tras tomarse una ducha rápida, se sienta en la mesa de la cocina para desayunar. Toi.Net[3], su compañero robot, le trae café caliente y un plato de fruta fresca. "Toi.Net, pásame el azúcar, por favor", dice Sam. El robot trae el azucarero desde el otro extremo de la mesa del desayuno. Hay un bote de azúcar en el armario de la cocina, pero Toi.Net sabe que es mucho más cómodo utilizar el azucarero.

"¿Alguna noticia interesante?", pregunta Sam. El robot deduce que debe encontrar noticias que correspondan a los temas de interés de Sam. Empieza por el fútbol.

Toi.Net: "El Mónaco ganó de local al Marsella por 3-1, es la primera vez que marca tres goles en un partido al Marsella en los últimos doce años. Un triplete de Diego Suárez".

Toi.Net: "El concurso de Eurovisión tuvo lugar en Liubliana; Polonia ganó con una canción sobre la amistad en las redes sociales".

2. El título de esta sección es una referencia a Samuel Butler, un novelista inglés del siglo XIX , autor de Erehwon, uno de los primeros libros que especula sobre la posibilidad de una inteligencia artificial desarrollada por la selección darwiniana y la reproducción entre máquinas.
3. Pronunciado 'tɔanət', como el nombre de la criada en "El inválido imaginario" de Molière.

Sam: "Por favor, no me vuelvas a molestar con este tipo de noticias, no me importa el concurso de Eurovisión".

Toi.Net: "De acuerdo. No lo haré".

Toi.Net: "La previsión meteorológica para París es de sol por la mañana, pero habrá algunas lluvias fuertes alrededor de la 1:00p.m. y por la tarde"

Toi.Net: "El Sr. Lamaison, candidato a la presidencia de la región suroeste, declaró que el nivel de desempleo alcanzó los 3,2 millones, su valor más alto desde 2004".

Sam: "¿Puedes comprobarlo? Creo recordar que el nivel era más alto a mediados de la década de 2010".

Toi.Net (después de dos segundos): "Tiene razón, llegó a 3,4 millones en 2015. Obtuve esa información de las estadísticas semánticas del Instituto Nacional de Estadística y de Estudios Económicos (INSEE)".

Al final del desayuno, Sam no se siente muy bien. Su pulsera conectada indica una presión arterial anormal y Toi.Net recibe la notificación. "¿Dónde ha dejado las pastillas?" le pregunta a Sam. "Las dejé en la mesita de noche, o quizá en el baño". Toi.Net le trae la caja de pastillas y Sam se recupera rápidamente.

Toi.Net: "Es hora de que vaya a trabajar. Como probablemente lloverá cuando vaya a dar un paseo por el parque después de comer, le he traído sus botines".

Un coche autónomo espera delante de la casa. Sam entra en el coche, que anuncia "Esta mañana voy a tomar un desvío por la A-4, ya que ha habido un accidente en su ruta habitual y hay 45 minutos de retraso por el taco".

Toi.Net es un robot bien educado. Sabe mucho sobre Sam, entiende sus peticiones, recuerda sus preferencias, puede encontrar objetos y actuar en consecuencia, se conecta a Internet y extrae información relevante, aprende de nuevas situaciones. Esto sólo ha sido posible gracias a los enormes avances de la inteligencia artificial: procesamiento y comprensión del habla (para entender las peticiones de Sam); visión y reconocimiento de objetos (para ubicar el azucarero en la mesa); planificación automática (para definir las secuencias de acciones correctas para llegar a una situación determinada, como entregar una caja de pastillas ubicada en otra habitación); representación del conocimiento (para identificar un triplete como una serie de tres goles hechos por el mismo jugador de fútbol)

razonamiento (para decidir escoger el azucarero en lugar del bote de azúcar del armario, o utilizar los datos de la previsión meteorológica para decidir qué par de zapatos debe llevar Sam); minería de datos (para extraer noticias relevantes de Internet, incluyendo la verificación de hechos en el caso de declaraciones políticas); su algoritmo de aprendizaje incremental hará que recuerde no mencionar los concursos de Eurovisión en el futuro; adapta continuamente sus interacciones con Sam construyendo el perfil de su propietario/a y detectando sus emociones.

Puestos a ser un poco provocadores, podemos decir que la inteligencia artificial no existe –pero, como es evidente, la potencia combinada de los datos, los algoritmos y los recursos informáticos disponibles abre enormes oportunidades en muchos ámbitos. Inria, con sus más de 200 equipos de proyectos, en su mayoría en colaboración con las principales universidades francesas, en ocho centros de investigación, trabaja en todas estas áreas científicas. Este libro blanco presenta nuestros puntos de vista sobre las principales tendencias y desafíos de la Inteligencia Artificial (IA) y cómo nuestros equipos están llevando a cabo de una manera activa la investigación científica, el desarrollo de software y la transferencia de tecnología en torno a estos desafíos clave para nuestra soberanía digital.

2

Una historia reciente de la IA

Está en boca de todos. Está en la televisión, la radio, los periódicos, las redes sociales. Vemos la IA en las películas, leemos sobre ella en las novelas de ciencia ficción. Nos encontramos con la IA cuando compramos nuestros billetes de tren por Internet o navegamos por nuestra red social favorita. Cuando escribimos su nombre en un buscador, el algoritmo encuentra hasta 16 millones de referencias. Tanto si nos fascina la mayoría del tiempo como si nos preocupa a veces, lo cierto es que nos obliga a cuestionarnos a nosotros mismos porque aún estamos lejos de saberlo todo sobre ella. Por todo ello, y esto es una realidad, la inteligencia artificial está entre nosotros. Los últimos años han sido un periodo en el que las empresas y los especialistas de diferentes campos (por ejemplo, Medicina, Biología, Astronomía, Humanidades Digitales) han mostrado un especial y pronunciado interés por los métodos de IA. Este interés suele ir acompañado de una visión clara de cómo la IA puede mejorar sus flujos de trabajo. El volumen de inversión tanto de las empresas privadas como de los gobiernos también supone un gran avance para la investigación en IA. Las principales empresas tecnológicas, pero también un número cada vez mayor de empresas industriales, participan ahora en la investigación sobre la IA y tienen previsto aumentar su inversión en el futuro, y muchos científicos especializados en IA dirigen ahora los laboratorios de investigación de estas y otras empresas.

La investigación en IA produjo importantes avances en la última década, en varias áreas. Los más mediáticos son los obtenidos en el aprendizaje automático, gracias en particular al desarrollo de arquitecturas de aprendizaje profundo, redes neuronales convolucionales multicapa que aprenden a partir de volúmenes masivos de datos y se entrenan en sistemas cómputo de alto rendimiento. Ya sea en la resolución de juegos, el reconocimiento de imágenes, el reconocimiento de voz y la traducción automática o la robótica, la inteligencia artificial se ha ido infiltrando en los últimos diez años en un gran número de aplicaciones industriales y de consumo que están revolucionando poco a poco nuestra relación con la tecnología.

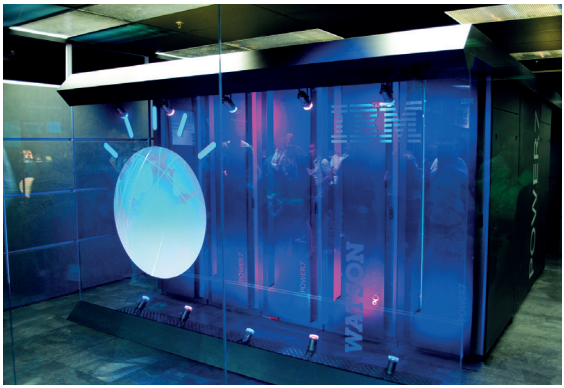


Figura 1: Computadora IBM Watson

En 2011, los científicos lograron desarrollar una inteligencia artificial capaz de procesar y comprender el lenguaje. La prueba se hizo pública cuando el software Watson de IBM ganó el famoso concurso Jeopardy. el principio del juego es proveer la pregunta a una respuesta dada tan rápido como sea posible El programa tenía que ser capaz de hacerlo igual de bien o incluso mejor para aspirar a vencer a los mejores: procesamiento del lenguaje, extracción de datos a gran velocidad, clasificación por nivel de probabilidad de las soluciones propuestas, todo ello con una alta dosis de computación intensiva. En la misma línea de Watson, el Proyecto Debater puede ahora realizar argumentaciones estructuradas discutiendo con expertos humanos, utilizando una mezcla de tecnologías (https://www.research.ibm.com/artificial-intelligence/project-debater/).

En otro escenario, la inteligencia artificial volvió a brillar en 2013 gracias a su capacidad para dominar siete videojuegos de Atari (computadora personal de las décadas 80 y 90). El aprendizaje por refuerzo desarrollado en el software de Google DeepMind permitió a su programa aprender a jugar a siete videojuegos y, sobre todo, a cómo ganar teniendo como única información los píxeles mostrados en la pantalla y la puntuación. El programa aprendió por sí mismo, a través de su propia experiencia, a mejorar continuamente y finalmente a ganar de forma sistemática. Desde entonces, el programa ha ganado una treintena de juegos Atari diferentes. Las proezas son aún más numerosas en los juegos de mesa de estrategia, sobre todo con AlphaGo, de Google Deepmind, que venció al campeón mundial de go en 2016 gracias a una combinación de aprendizaje profundo y aprendizaje por refuerzo, combinado con múltiples entrenamientos con humanos, otras computadoras y consigo mismo. El algoritmo fue mejorado en las siguientes versiones: en 2017, AlphaZero alcanzó un nuevo nivel al entrenar solo contra sí mismo, es decir, por autoaprendizaje. En un tablero de go, ajedrez o damas, ambos jugadores conocen la situación exacta de la partida en todo momento. Las estrategias son calculables hasta cierto punto: según las jugadas posibles, hay soluciones óptimas y un programa bien diseñado es capaz de identificarlas. Pero, ¿qué pasa con un juego a base de faroles e información oculta? En 2017, Tuomas Sandholm, de la Universidad Carnegie-Mellon, presentó el programa Libratus, que aplastó mediante aprendizaje a cuatro de los mejores jugadores en una competición de póquer, véase https://www.cs.cmu.edu/~noamb/papers/17-IJCAI-Libratus.pdf. En consecuencia, la resolución de problemas con incógnitas por parte de la IA podría beneficiar a muchos ámbitos, como las finanzas, la salud, la ciberseguridad o la defensa. Sin embargo, hay que tener en cuenta que incluso los juegos de mesa con información incompleta que la IA ha "resuelto" recientemente (el póquer, como se ha descrito anteriormente, StarCraft de DeepMind, Dota2 de Open AI) tienen lugar en un universo conocido: las acciones del oponente son desconocidas, pero su distribución de probabilidad es conocida, y el conjunto de acciones posibles es

finito, aunque sea enorme. Por el contrario, el mundo real implica generalmente un número infinito de situaciones posibles, lo que hace que la generalización sea mucho más difícil.

Entre los logros más destacados de los últimos tiempos se encuentran los realizados en el desarrollo de los vehículos autónomos y conectados, que son objeto de colosales inversiones por parte de los fabricantes de automóviles que van haciendo realidad paulatinamente el mito del vehículo totalmente autónomo con un conductor totalmente pasivo que se convertiría así en un pasajero. Más allá del marketing comercial de los fabricantes, los avances son bastante reales y también anuncian un fuerte desarrollo de estas tecnologías, pero en una escala de tiempo muy diferente. Los vehículos autónomos han recorrido millones de kilómetros con sólo unos pocos incidentes importantes. En pocos años, la IA se ha consolidado en todos los ámbitos de los vehículos autónomos conectados (CAV, por su sigla en inglés), desde la percepción hasta el control, pasando por la decisión, la interacción y la supervisión. Esto abrió el camino a soluciones que antes no eran efectivas y abrió también nuevos retos de investigación (por ejemplo, la conducción de extremo a extremo). El aprendizaje profundo, en particular, se convirtió en una herramienta común y versátil, fácil de implementar y desplegar.

Esto ha motivado el desarrollo acelerado de hardware y arquitecturas específicas, como las tarjetas de procesamiento dedicadas que la industria del automóvil integra a bordo de vehículos autónomos reales y plataformas prototipo.

En su libro blanco, Autonomous and Connected Vehicles: Current Challenges and Research Paths, publicado en mayo de 2018, Inria advierte, no obstante, de los límites de la implementación a gran escala: ***"Los primeros sistemas automatizados de transporte, en sitios privados o de acceso controlado, deberían aparecer a partir de 2025. En ese momento, los vehículos autónomos también deberían empezar a circular por las autopistas, siempre y cuando la infraestructura se haya adaptado (por ejemplo, en carriles exclusivos). No será hasta 2040 cuando veamos coches completamente autónomos, en zonas peri-urbanas y en pruebas en las ciudades",*** afirma Fawzi Nashashibi, jefe del equipo-proyecto RITS en Inria y principal autor del libro blanco. ***"Pero la madurez de las tecnologías no es el único obstáculo para la implementación de estos vehículos, que dependerá en gran medida de las decisiones políticas (inversiones, normativas, etc.) y de las estrategias de ordenación del territorio",*** prosigue.

En el ámbito de la salud y la medicina, véase, por ejemplo, el libro de Eric Topol "Deep Medicine", que muestra docenas de aplicaciones del aprendizaje profundo en casi todos los aspectos de la salud, desde la radiografía hasta el

diseño de dietas y la rehabilitación mental. Un logro clave en los últimos tres años es el rendimiento de Deepmind en CASP (Critical Assessment of Structure Prediction) con AlphaFold, un método que superó significativamente a todos los contendientes en la tarea de predecir la estructura tridimensional de proteínas a partir de su secuencia.Estos resultados abren una nueva era: sería posible obtener las estructuras de alta resolución para la gran mayoría de proteínas de las cuales sólo se conoce su secuencia. Otro logro clave es la estandarización de los conocimientos, en particular de la regulación biológica, que es muy compleja de unificar (formato BioPAX) y las numerosas bases de conocimiento disponibles (Reactome, Rhea, pathwaysCommnons, etc.). Mencionemos también el interés y la energía mostrados por ciertos médicos, en particular los radiólogos, en las herramientas relacionadas con el diagnóstico y el pronóstico automatizados, en particular en el campo de la oncología. En 2018, la FDA permitió la comercialización de IDx-DR (https://www.eyediagnosis.co/), el primer dispositivo médico que utiliza la IA para detectar algo más que un nivel leve de retinopatía diabética en el ojo de adultos con diabetes (https://doi.org/10.1038/s41433-019-0566-0).

En el sector de la aviación, las Fuerzas Aéreas de Estados Unidos han desarrollado, en colaboración con la empresa Psibernetix, un sistema de IA capaz de vencer a los mejores pilotos humanos en combate aéreo[4]. Para lograrlo, Psibernetix combina algoritmos de lógica difusa y un algoritmo genético, es decir, un algoritmo que se inspira en los mecanismos de la evolución biológica. Esto permite a la IA centrarse en lo esencial y desglosar sus decisiones en los pasos que deben resolverse para lograr su objetivo.

Al mismo tiempo, la robótica también se está beneficiando de muchos de los nuevos avances tecnológicos, especialmente gracias al Darpa Robotics Challenge, organizado de 2012 a 2015 por la Agencia de Investigación Avanzada del Departamento de Defensa de Estados Unidos (https://www.darpa.mil/program/darpa-robotics-challenge). Este concurso demostró que era posible desarrollar robots terrestres semi autónomos capaces de realizar tareas complejas en entornos peligrosos y degradados: conducir vehículos, manipular válvulas, progresar en entornos de riesgo. Estos avances apuntan a una multitud de aplicaciones, ya sean militares, industriales, médicas, domésticas o recreativas.

4. https://magazine.uc.edu/editors_picks/recent_features/alpha.html

Otros ejemplos notables son:

→ **Descripción automática del contenido de una imagen** ("una imagen vale más que mil palabras"), también de Google (http://googleresearch.blogspot.fr/2014/11/a-picture-isworth-thousand-coherent.html)

→ **Los resultados del Desafío de Reconocimiento Visual a gran Escala de Imagenet de 2012**, ganado por una red neuronal convolucional profunda desarrollada por la Universidad de Toronto (http://image-net.org/challenges/LSVRC/2012/results.html)

→ **La calidad de los sistemas de reconocimiento facial como el de Facebook,** https://www.newscientist.com/article/dn27761-facebook-can-recognise-youin-photos-even-if-youre-not-looking - .VYkVxFzjZ5g

→ **Flash Fill,** una función automática de Excel, adivina una operación repetitiva y la completa (en programación por ejemplo). Sumit Gulwani: Automating string processing in spreadsheets using input-output examples. POPL 2011: 317-330.

→ **PWC-Net de Nvidia** ganó la competición de etiquetado con movimiento óptico de 2017 usando los conjuntos de datos de MPI Sintel y KITTI 2015, combinando modelos de aprendizaje profundo y conocimiento del dominio." https://arxiv.org/abs/1709.02371

→ El procesamiento del habla es ahora una característica estándar de los teléfonos inteligentes y las tabletas con asistentes virtuales como Siri de Apple, Alexa de Amazon, Cortana de Microsoft y otros. Google Meet transcribe el diálogo de los participantes en una reunión en tiempo real. Los auriculares Ambassador de Waverly Labs traducen las conversaciones en diferentes idiomas; la traducción simultánea está presente en Skype de Microsoft desde hace muchos años.

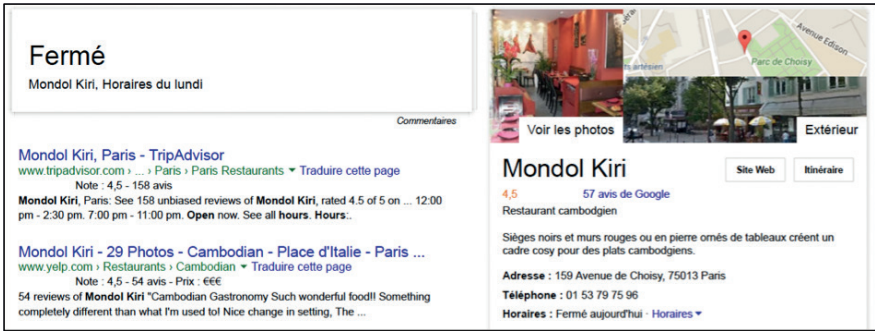


Figura 2: Información semántica añadida a los resultados del motor de búsqueda de Google

También cabe mencionar los resultados obtenidos en la representación del conocimiento y el razonamiento, las ontologías y otras tecnologías para la web semántica y para los datos enlazados:

→ Google Knowledge Graph mejora los resultados de búsqueda mostrando datos estructurados sobre los términos o frases de búsqueda solicitados. En el ámbito de la web semántica, se observa el aumento de la capacidad para responder a peticiones específicas como "maridos de las hijas de Marie Curie" y para interpretar los datos RDF que se pueden encontrar en la web.

Figura 3: Procesamiento semántico en la web

→ Schema.org[5] contiene millones de tripletas RDF (Resource Description Framework) que describen hechos conocidos: los motores de búsqueda pueden utilizar estos datos para proporcionar información estructurada si así se requiere.

→ El protocolo OpenGraph -que utiliza RDFa- es utilizado por Facebook para permitir que cualquier página web se convierta en un objeto enriquecido en un gráfico social.

Otra tendencia importante es la reciente liberación de varias tecnologías que antes eran privadas, para que la comunidad de investigadores de IA se beneficie de ellas, pero también para que contribuya con características adicionales. Ni que decir tiene que esta apertura es también una estrategia de las Big Tech para construir y organizar comunidades de conocimiento y de usuarios centradas en sus tecnologías. Algunos ejemplos son:

→ Los servicios de computación cognitiva de IBM para Watson, disponibles a través de sus Interfaces de Programación de Aplicaciones, ofrecen hasta 20 tecnologías diferentes, como voz a texto y texto a voz, identificación y vinculación de conceptos, reconocimiento visual y muchas otras: https://www.ibm.com/watson

5. https://schema.org/

→ **TensorFlow de Google** es la biblioteca de software de código abierto más popular para el aprendizaje automático; https://www.tensorflow.org/. Se puede encontrar una buena muestra de las principales plataformas de código abierto para el aprendizaje automático en http://aiindex.org

→ Facebook hizo público su diseño de hardware Big Sur para ejecutar grandes redes neuronales de aprendizaje profundo en las GPU: https://ai.facebook.com/blog/the-nextstep-in-facebooks-ai-hardware-infrastructure/

Además de estas herramientas anteriormente protegidas, algunas bibliotecas se desarrollaron de forma nativa como software de código abierto. Este es el caso, por ejemplo, de la biblioteca Scikit-learn (véase la Sección 5.2.5), un activo estratégico en el compromiso de Inria en este campo.

Por último, para concluir este capítulo veamos algunos logros científicos de la IA:

Aprendizaje automático:

→ Cuestionamiento empírico de conceptos estadísticos teóricos que parecían firmemente establecidos. La teoría había sugerido claramente que había que evitar el régimen sobre parametrizado para evitar el peligro del sobreaprendizaje. Numerosos experimentos con redes neuronales han demostrado que el comportamiento en el régimen sobre parametrizado es mucho más estable de lo esperado, y han generado un entusiasmo renovado para comprender teóricamente los fenómenos implicados.

→ Los enfoques de la física estadística se han utilizado para determinar los límites fundamentales de viabilidad de varios problemas de aprendizaje, así como los algoritmos eficientes asociados a ello.

→ Se desarrollaron integraciones (representaciones de datos de baja dimensión) y se utilizaron como entrada de arquitecturas de aprendizaje profundo para casi todas las representaciones, por ejemplo, word2vec para el lenguaje natural, graph2vec para los grafos, math2vec para las matemáticas, bio2vec para los datos biológicos, etc.

→ El alineamiento de grafos o de nubes de puntos ha avanzado mucho tanto en la teoría como en la práctica, dando, por ejemplo, resultados sorprendentes en la capacidad de construir diccionarios bilingües de forma no estructurada.

➔ Los transformadores que utilizan redes neuronales profundas de gran tamaño y mecanismos de atención han llevado la vanguardia del procesamiento del lenguaje natural a nuevos horizontes. Los sistemas basados en transformadores son capaces de entablar conversaciones sobre cualquier tema con usuarios humanos.

➔ Los métodos híbridos que mezclan expresividad lógica, incertidumbre y el rendimiento de las redes neuronales están empezando a producir resultados interesantes, véase, por ejemplo, https://arxiv.org/pdf/1805.10872.pdf de De Raedt et al.; También es el caso de los trabajos que mezclan métodos simbólicos y numéricos para resolver problemas de forma diferente a como se ha hecho durante años, por ejemplo, "Anytime discovery of a diverse set of patterns with Monte Carlo tree search". https://arxiv.org/abs/1609.08827 Véase también el trabajo de Serafini y d'Avila Garcez sobre "Logic tensor networks" que conecta las redes neuronales profundas con las restricciones expresadas en lógica. https://arxiv.org/abs/1606.04422

Procesamiento de imágenes y vídeos:

➔ Desde la revelación de los rendimientos del aprendizaje profundo en la campaña Imagenet de 2012, la calidad y la precisión de la detección y seguimiento de objetos (por ejemplo, las personas con su postura) experimentaron avances significativos. Las aplicaciones son ahora posibles, aunque siguen existiendo muchos retos.

Procesamiento del lenguaje natural (NLP)

➔ Los modelos neuronales de la NLP (traducción automática, generación de textos, minería de datos) han experimentado un progreso espectacular con, por un lado, nuevas arquitecturas (redes de transformadores que utilizan mecanismos de atención) y, por otro, la idea de pre-entrenar representaciones de palabras o frases mediante algoritmos de aprendizaje no supervisado que luego pueden utilizarse de forma provechosa en tareas específicas con muy pocos datos supervisados.

➔ Se han obtenido resultados espectaculares en la traducción no supervisada, así como en el campo de las representaciones multilingües y en el reconocimiento automático del habla, con una reducción de 100 veces en los datos etiquetados (¡10h en lugar de 1000h!), utilizando un pre-entrenamiento no supervisado sobre audio sin etiquetar[6].

6. https://arxiv.org/abs/2006.11477.

Redes Generativas Antagónicas (GANs):

➔ Los resultados obtenidos por las redes neuronales generativas antagónicas (GAN) son particularmente impresionantes. Son capaces de generar imágenes naturales realistas a partir de ruido aleatorio. Aunque la comprensión de estos modelos es todavía limitada, han mejorado significativamente nuestra capacidad para extraer muestras de distribuciones de datos especialmente complejas. A partir de distribuciones aleatorias, las GAN pueden producir música nueva, generar deepfakes realistas, escribir frases de texto comprensibles, y cosas por el estilo.

Optimización:

➔ Los problemas de optimización que parecían imposibles hace unos años pueden resolverse ahora con métodos casi genéricos. La combinación de aprendizaje automático y optimización abre vías para la resolución de problemas complejos en el diseño, el funcionamiento y la supervisión de sistemas industriales. Para ello, proliferan las herramientas y bibliotecas para la IA que pueden acoplarse fácilmente a los métodos de optimización y solucionadores.

Representación del conocimiento:

➔ El creciente interés por combinar grafos de conocimiento e incrustaciones de grafos para realizar un aprendizaje automático (semántico) basado en grafos

➔ Nuevas direcciones como la IA de proximidad basada en la web. https://www.w3.org/wiki/Networks/Edge_computing

Por supuesto, todos estos resultados tienen limitaciones científicas y tecnológicas; los retos pertinentes se presentan más adelante en el capítulo 5.

Por otro lado, estos logros positivos se han visto equilibrados por ciertas preocupaciones sobre los peligros de la IA expresadas por científicos muy reconocidos, y más globalmente por muchas partes interesadas en la IA, que es el tema de la siguiente sección.

3

Debates sobre la IA

Los debates sobre la IA empezaron realmente en el siglo XX –por ejemplo, recordemos las Leyes de la Robótica de Isaac Asimov–, pero se han intensificado debido a los recientes avances logrados por los sistemas de IA, como se muestra arriba. La teoría de la singularidad tecnológica afirma que una nueva era de las máquinas que dominarán el mundo comenzará una vez que los sistemas de IA se vuelvan súper inteligentes:

"La singularidad tecnológica es un hecho hipotético relacionado con la llegada de una auténtica inteligencia artificial fuerte. La singularidad tecnológica implica que un equipo de cómputo, red informática o un robot podrían ser capaces de automejorarse recursivamente (rediseño de sí mismo), o en el diseño y construcción de computadoras o robots mejores que él mismo. Se dice que las repeticiones de este ciclo probablemente darían lugar a un efecto fuera de control —una explosión de inteligencia— donde las máquinas inteligentes podrían diseñar generaciones de máquinas sucesivamente más potentes. La creación de inteligencia sería muy superior al control y la capacidad intelectual humana. Dado que las capacidades de una superinteligencia de este tipo pueden ser imposibles de comprender para un ser humano, la singularidad tecnológica es el punto a partir del cual los acontecimientos pueden resultar imprevisibles o incluso insondables para la inteligencia humana" (Wikipedia).

Los defensores de la singularidad tecnológica están próximos al movimiento transhumanista, que aspira a mejorar las capacidades físicas e intelectuales de los seres humanos con nuevas tecnologías. La singularidad sería el momento en el que la naturaleza de los seres humanos cambiaría por completo, lo que se percibe o bien como un hecho deseable, o bien como un peligro para la humanidad.

Un resultado importante del debate sobre los peligros de la IA ha sido la controversia sobre las armas autónomas y los robots asesinos, apoyada por una carta abierta publicada en la apertura de la conferencia del IJCAI en 2015[7]. La carta, que pide la prohibición de este tipo de armas capaces de operar sin ningún control humano, ha sido firmada por miles de personas, entre ellas Stephen Hawking, Elon Musk, Steve Wozniak y una serie de destacados investigadores de IA, incluidos algunos de Inria, colaboradores de este documento. Véase también el vídeo "Slaughterbots" de Stuart Russell[8].

Otros peligros y amenazas que se han debatido en la comunidad son las consecuencias financieras en los mercados bursátiles de la negociación de alta

7. Véase http://futureoflife.org/open-letter-autonomous-weapons/
8. https://www.youtube.com/watch?v=HipTO_7mUOw

frecuencia, que ahora representa la gran mayoría de las órdenes realizadas, en las que los programas informáticos supuestamente inteligentes (que en realidad se basan en la toma de decisiones estadísticas que no pueden ser consideradas como IA) operan a un ritmo frenético que conduce a posibles colapsos del mercado, como en el caso del Flash Crash de 2010; las repercusiones de la minería de big data en la privacidad, con sistemas de minería capaces de divulgar información privada de las personas al establecer vínculos entre sus operaciones en línea o sus registros en bancos de datos; y, por supuesto, el peligro de desempleo provocado por la progresiva sustitución de la mano de obra por máquinas.

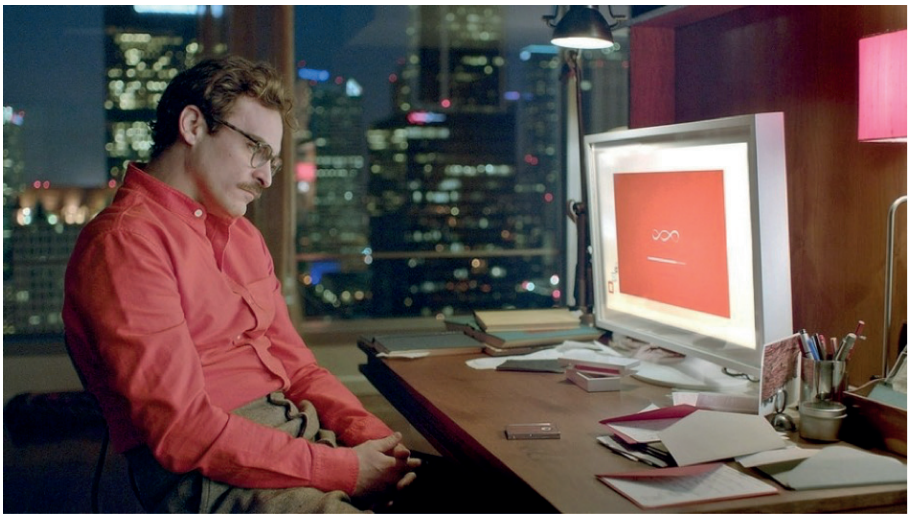


Figura 4: En la película "Her" de Spike Jonze, un hombre se enamora de su sistema operativo inteligente

Cuanto más desarrollemos la inteligencia artificial, mayor será el riesgo de desarrollar sólo determinadas capacidades de la inteligencia (por ejemplo, la optimización y la minería por aprendizaje) en detrimento de otras para las que el retorno de la inversión puede no ser inmediato o incluso no ser una preocupación para el creador del agente (por ejemplo, moral, respeto, ética, etc.). Cuando se utiliza a gran escala, la inteligencia artificial puede plantear muchos riesgos y desafíos para los seres humanos, especialmente si las inteligencias artificiales no se diseñan y supervisan de forma que respeten y protejan a los seres humanos. Si, por ejemplo, la optimización y el rendimiento son el único objetivo de su inteligencia, esto puede conducir a desastres a gran escala en los que los usuarios son instrumentalizados, abusados, manipulados, etc. por agentes artificiales imparables y desvergonzados. La investigación sobre IA debe ser exhaustiva e incluir todo lo que hace que los comportamientos sean inteligentes, no sólo los

"aspectos más razonables". Esto va más allá de las cuestiones puramente científicas y tecnológicas, pues lleva a cuestiones de gobernanza y regulación.

Dietterich y Horvitz publicaron una interesante respuesta a algunas de estas preguntas[9]. En su breve artículo, los autores reconocen que la comunidad de investigadores de IA debería poner solo una atención moderada al riesgo de pérdida de control por parte de los humanos, porque no es crítico en un futuro próximo, y en su lugar recomiendan que se preste más atención a los cinco riesgos a corto plazo a los que se enfrentan los sistemas basados en IA, a saber: los fallos en el software; los ciberataques; "El aprendiz de brujo", es decir, hacer que los sistemas de IA entiendan lo que quieren las personas en lugar de interpretar literalmente sus órdenes; la "autonomía compartida", es decir, la cooperación fluida de los sistemas de IA con los usuarios, para que éstos puedan recuperar siempre el control en caso necesario; y las repercusiones socioeconómicas de la IA: en otras palabras, la IA debe beneficiar a la sociedad en su conjunto y no sólo a unos pocos privilegiados.

En los últimos años, los debates se han centrado en una serie de cuestiones en torno a la noción de una IA responsable y digna de confianza, que podemos resumir como sigue:

→ **Confianza:** Nuestras interacciones con el mundo y entre nosotros se canalizan cada vez más a través de herramientas de IA. ¿Cómo garantizar los requisitos de seguridad de las aplicaciones más importantes, la seguridad y la confidencialidad de los medios de comunicación y procesamiento? ¿Qué técnicas y normas de validación, certificación y auditoría de las herramientas de IA deben desarrollarse para generar confianza en la IA?

→ **Gobernanza de datos:** El vínculo entre los datos, la información, el conocimiento y las acciones está cada vez más automatizado y es más eficiente. ¿Qué normas de gobernanza de datos de todo tipo, personales, metadatos y datos agregados a varios niveles, son necesarias? ¿Qué instrumentos harían posible su cumplimiento? ¿Cómo garantizar la trazabilidad de los datos desde los productores hasta los consumidores?

→ **Empleo:** La automatización acelerada de las actividades físicas y cognitivas tiene fuertes repercusiones económicas y sociales. ¿Cuáles son sus efectos sobre la transformación y la división social del trabajo? ¿Cuáles son las repercusiones

9. Dietterich, Thomas G. y Horvitz, Eric J., Rise of Concerns about AI: Reflections and Directions, Communications of the ACM, October 2015 Vol. 58 no. 10, pp. 38-40

en los intercambios económicos? ¿Qué medidas proactivas y de adaptación serán necesarias? ¿Es esta situación diferente a la de las anteriores revoluciones industriales?

→ **Supervisión humana:** Cada vez delegamos más decisiones personales y profesionales en las PDA. ¿Cómo podemos beneficiarnos de ello sin el riesgo de alienación y manipulación? ¿Cómo hacer que los algoritmos sean inteligibles, que produzcan explicaciones claras y que sus funciones de evaluación se correspondan con nuestros valores y criterios? ¿Cómo podemos anticiparnos y restablecer el control humano cuando el contexto está fuera del alcance de esta delegación?

→ **Sesgos:** Nuestros algoritmos no son neutrales; se basan en las suposiciones y sesgos implícitos, a menudo no intencionados, de sus diseñadores o presentes en los datos utilizados para el aprendizaje. ¿Cómo identificar y superar estos sesgos? ¿Cómo diseñar sistemas de IA que respeten los valores humanos esenciales, y que no acentúen las desigualdades?

→ **Privacidad y seguridad:** Las aplicaciones de IA pueden plantear problemas de privacidad, por ejemplo en el caso del reconocimiento facial, una tecnología útil para facilitar el acceso a los servicios digitales, pero una tecnología cuestionable cuando se generaliza su uso. ¿Cómo podemos diseñar sistemas de IA que no vulneren de forma innecesaria las restricciones de privacidad? ¿Cómo podemos garantizar la seguridad y la fiabilidad de las aplicaciones de IA que pueden ser objeto de ataques de adversarios?

→ **Sostenibilidad:** Los sistemas de aprendizaje automático utilizan una cantidad de potencia y energía informática cada vez mayor, debido al volumen de datos de entrada y al número de parámetros que deben ser optimizados. ¿Cómo podemos construir sistemas de IA cada vez más sofisticados utilizando recursos limitados?

Evitar los riesgos es necesario, pero no suficiente para poner la IA al servicio de la humanidad. ¿Cómo podemos dedicar una parte sustancial de nuestros recursos de investigación y desarrollo a los grandes retos de nuestro tiempo (clima, medio ambiente, salud, educación) y, en un sentido más amplio, a los objetivos de desarrollo sostenible de la ONU?

Estas y otras cuestiones deben ser objeto de reflexiones ciudadanas y políticas, experimentos controlados, observatorios de usos y opciones sociales. Se han documentado en varios informes que ofrecen recomendaciones, directrices y

principios para la IA, como la Declaración de Montreal para una IA responsable[10], las Recomendaciones de la OCDE sobre Inteligencia Artificial[11], las Directrices Éticas para una Inteligencia Artificial Fiable del Grupo de Expertos de Alto Nivel de la Comisión Europea[12] y muchos otros, como la UNESCO, el Consejo de Europa, gobiernos, empresas privadas, ONG, etc. En total, hay más de cien documentos de este tipo al momento de redactar este libro blanco.

Inria es consciente de estos debates y, como instituto de investigación dedicado a las ciencias digitales y a la transferencia tecnológica, trabaja por el bienestar de todos, plenamente consciente de sus responsabilidades con la sociedad. Informar a la sociedad y a los órganos de gobierno sobre el potencial y los riesgos de la ciencia y la tecnología digitales forma parte de la misión de Inria.

Inria inició una reflexión sobre la ética mucho antes de que las amenazas de la IA suscitaran debates en la comunidad científica. En los últimos años, Inria:

→ Contribuyó a la creación del CERNA de Allistene[13], un grupo de reflexión que estudia las cuestiones éticas derivadas de la investigación sobre ciencia y tecnologías digitales; los dos primeros informes de recomendaciones publicados por el CERNA trataban sobre la investigación en robótica y sobre las mejores prácticas para el aprendizaje automático;

→ Crear un órgano encargado de evaluar las cuestiones jurídicas y éticas de un proyecto de investigación caso por caso: el Comité Operativo de Evaluación de Riesgos Jurídicos y Éticos (COERLE), compuesto por científicos de Inria y colaboradores externos; la misión del COERLE es ayudar a identificar los riesgos y determinar si es necesaria la supervisión de un determinado proyecto de investigación;

→ Participó intensamente en la creación de nuestro comité nacional de ética de las tecnologías digitales[14];

→ Se encargó de la coordinación del componente de investigación de la estrategia de IA de nuestro país (véase el capítulo 4);

→ El gobierno francés le pidió que organizara el Foro Global sobre Inteligencia

10. https://www.montrealdeclaration-responsibleai.com/
11. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449
12. Grupo de expertos de alto nivel de la Comisión Europea (2018). Ethics Guidelines for Trustworthy AI
13. Comisión de reflexión sobre la ética de la investigación en ciencias y tecnologías digitales de la Alianza de Ciencias y Tecnologías Digitales: https://www.allistene.fr/cerna/
14. https://www.allistene.fr/tag/cerna/

Artificial para la Humanidad, un coloquio que reunió a los principales expertos mundiales en IA y sus consecuencias sociales, a finales de 2019[15], como precursor de la GPAI (véase más adelante);

→ Se le encomendó la responsabilidad del Centro de Expertos de París de la Asociación Mundial sobre Inteligencia Artificial, una iniciativa internacional y de múltiples sectores, para orientar en el desarrollo y el uso responsables de la inteligencia artificial en consonancia con los derechos humanos, las libertades fundamentales y los valores democráticos compartidos, puesta en marcha por catorce países y la Unión Europea en junio de 2020.

Además, Inria alienta a sus investigadores a participar en los debates sociales cuando son convocados por la prensa y los medios de comunicación para hablar sobre cuestiones éticas como las que plantean la robótica, el aprendizaje profundo, la minería de datos y los sistemas autónomos. Inria también contribuye a la educación del público invirtiendo en el desarrollo de cursos en línea de acceso universal y abierto (Massive Open Online Courses, o MOOC por sus siglas en inglés) sobre IA y sobre algunos de sus subdominios ("L'intelligence artificielle avec intelligence"[16], "Web sémantique et web de données"[17], "Binaural hearing for robots"[18]) y, de forma más general, desempeñando un papel activo en las iniciativas educativas para las ciencias digitales.

Dicho esto, veamos ahora los retos científicos y tecnológicos de la investigación en IA, y cómo contribuye Inria a afrontarlos: este será el tema de la siguiente sección.

15. https://www.youtube.com/playlist?list=PLJ1qHZpFsMsTXDBLLWIkAUXQG_d5Ru3CT
16. https://www.fun-mooc.fr/courses/course-v1:inria+41021+session01/about
17. https://www.fun-mooc.fr/courses/course-v1:inria+41002+autogestión/sobre
18. https://www.fun-mooc.fr/courses/course-v1:inria+41004+archiveouvert/about

4

Inria como parte de la estrategia nacional de IA

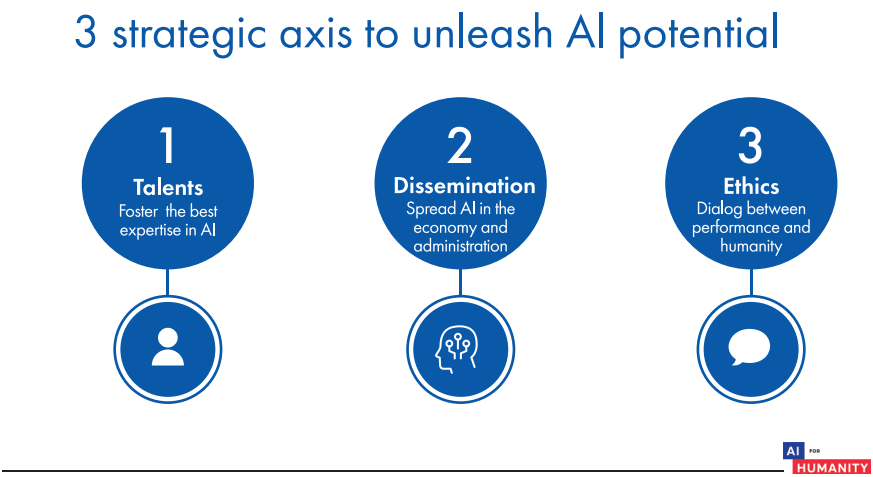
AI FOR HUMANITY: EL PROGRAMA NACIONAL DE INVESTIGACIÓN EN IA

En la jornada de clausura del debate "AI for Humanity" (IA para la Humanidad) celebrado en París el 29 de marzo de 2018, el Presidente de la República Francesa presentó una ambiciosa estrategia para la Inteligencia Artificial (IA) y lanzó la Estrategia Nacional de IA (https://www.aiforhumanity.fr/en/).

La Estrategia Nacional de Inteligencia Artificial pretende convertir a Francia en líder de la IA, un sector dominado actualmente por Estados Unidos y China, y por países emergentes en este sector como Israel, Canadá y el Reino Unido.

Las prioridades que el Presidente de la República estableció son la investigación, los datos abiertos y las cuestiones éticas o sociales. Estas medidas se desprenden del informe elaborado por el matemático y diputado Cédric Villani, que realizó audiencias con más de 300 expertos de todo el mundo. Para concluir este proyecto, Cédric Villani trabajó con Marc Schoenauer, director de investigación y jefe del equipo-proyecto TAU en el centro de investigación Inria Saclay - Île-de-France.

Esta Estrategia Nacional de IA, con un presupuesto de 1.500 millones de euros de dinero público durante cinco años, reúne tres ejes: (i) lograr el mejor nivel de investigación para la IA, formando y atrayendo a los mejores talentos del mundo en este campo; (ii) propagar la IA en la economía y la sociedad a través de spin-offs y asociaciones público-privadas y el intercambio de datos; (iii) establecer un marco ético para la IA. Ya se han tomado muchas medidas en estos tres ámbitos.



Como parte del Plan de IA para la Humanidad, se encomendó a Inria la coordinación del Programa Nacional de Investigación en IA. El plan de investigación interactúa con cada uno de los tres ejes mencionados.

La reunión inicial del eje de investigación se celebró en Toulouse el 28 de noviembre de 2018. El objetivo del **Programa Nacional de Investigación en IA** (https://www.inria.fr/en/ai-mission-national-artificial-intelligence-researchprogram) tiene un doble objetivo: situar a Francia de forma duradera entre los cinco primeros países en IA y convertir a Francia en un líder europeo en investigación en IA.

Para ello, se llevarán a cabo varias acciones en una primera etapa que durará desde finales de 2018 hasta 2022:

- Crear una red nacional de investigación en IA coordinada por Inria;
- Poner en marcha 4 Institutos Interdisciplinarios de Inteligencia Artificial;
- Promover programas de atracción y apoyo al talento en todo el país;
- Contribuir al desarrollo de un programa específico de formación en IA;
- Aumentar los recursos informáticos dedicados a la IA y facilitar el acceso a las infraestructuras;
- Impulsar las asociaciones público-privadas;
- Impulsar la investigación en IA a través de las convocatorias de la Agencia Nacional de Investigación Francesa (ANR, Agence Nationale de la Recherche)
- Reforzar la cooperación bilateral, europea e internacional;

El pilar de investigación también está en contacto con las iniciativas de innovación en IA, en particular con los Grandes Retos del Consejo de Innovación (https://www.gouvernement.fr/decouvrir-lesgrands-defis).

5

Los retos de la IA y las contribuciones de Inria

El enfoque de Inria consiste en combinar simultáneamente dos cometidos: comprender los sistemas que están operando en el mundo (desde los sociales hasta los tecnológicos), y los problemas que surgen de sus interacciones; y actuar sobre ellos para encontrar soluciones al proporcionar modelos numéricos, algoritmos, software y tecnologías. Esto implica desarrollar una descripción precisa, por ejemplo formal o aprendida a partir de datos, herramientas adecuadas para razonar sobre ella o manipularla, así como proponer soluciones innovadoras y eficaces. Esta visión se ha desarrollado a lo largo de los 50 años de existencia del instituto, favorecida por una organización que no separa la teoría de la práctica, ni las matemáticas de la informática, sino que reúne los conocimientos necesarios en equipos de investigación consolidados, sobre la base de proyectos de investigación específicos.

La noción de "ciencias digitales" no está definida de forma absoluta, pero podemos acercarnos a ella a través del doble objetivo expuesto anteriormente, comprender el mundo y luego actuar sobre él. El desarrollo del "pensamiento computacional" requiere la capacidad de definir, organizar y manipular los elementos que constituyen el núcleo de las ciencias digitales: Modelos, Datos y Lenguajes. El desarrollo de técnicas y soluciones para el mundo digital exige la investigación en diversos ámbitos, que suelen mezclar modelos matemáticos, avances algorítmicos y sistemas. Por lo tanto, identificamos las siguientes ramas en la investigación relevante para Inria:

- → **Algoritmos y programación,**
- → **Ciencia de datos e ingeniería del conocimiento,**
- → **Modelización y simulación,**
- → **Optimización y control.**
- → **Arquitecturas, sistemas y redes,**
- → **Seguridad y confidencialidad,**
- → **Interacción y multimedia,**
- → **Inteligencia artificial y sistemas autónomos.**

Como toda clasificación, esta presentación es en parte arbitraria y no expone las numerosas interacciones entre los distintos ámbitos. Por ejemplo, los estudios sobre

redes también implican desarrollos de algoritmos novedosos, y la inteligencia artificial es de naturaleza muy transversal, con fuertes vínculos con la ciencia de datos. Evidentemente, cada una de estas ramas es un área de investigación muy activa en la actualidad. Inria ha invertido en estos ámbitos mediante la creación de equipos de proyecto específicos y la construcción de una sólida experiencia en muchos de ellos. Cada una de estas ramas se considera importante para el instituto.

La IA es un campo muy amplio; cualquier intento de estructurarlo en subcampos puede generar debates. Utilizaremos la jerarquía de palabras clave propuesta por la comunidad de jefes de equipo de Inria para identificar mejor sus contribuciones a las ciencias digitales en general. En esta jerarquía, la Inteligencia Artificial es una palabra clave de primer nivel con ocho subdominios, algunos de ellos específicos y otros que se refieren a otras secciones de la jerarquía: véase el cuadro siguiente:

- **Conocimiento**
  - Bases de conocimiento
  - Extracción de conocimiento y limpieza
  - Inferencia
  - Web semántica
  - Ontologías
- **Aprendizaje automático**
  - Aprendizaje supervisado
  - Aprendizaje no supervisado
  - Aprendizaje secuencial y por refuerzo
  - Optimización para el aprendizaje
  - Métodos bayesianos
  - Redes neuronales
  - Métodos Kernel
  - Aprendizaje profundo
  - Minería de datos
  - Análisis masivo de datos
- **Procesamiento del lenguaje natural**

- **Procesamiento de señal (habla, visión)**
  - **Habla**
  - **Visión**
    - Reconocimiento de objetos
    - Reconocimiento de actividad
    - Búsqueda en bancos de imágenes y vídeos
    - Reconstrucción 3D y espacio-temporal
    - Seguimiento de objetos y análisis de movimiento
    - Localización de objetos
    - Control robótico basado en visión

- **Robótica (incluidos los vehículos autónomos)**
  - Diseño
  - Percepción
  - Decisión
  - Acción
  - Interacción con los robots (entorno/humanos/robots)
  - Flotas de robots
  - Aprendizaje de robots
  - Cognición para la robótica y los sistemas

- **Neurociencias, ciencias cognitivas**
  - Comprensión y simulación del cerebro y del sistema nervioso
  - Ciencias cognitivas

- **Algorítmica de la IA**
  - Programación lógica y ASP (Answer set programming en su denominación inglesa)
  - Deducción y prueba
  - Teorías de la acción contextual (SAT; Situational Action Theory)
  - Razonamiento causal, temporal y sobre incertidumbre
  - Programación con restricciones
  - Búsqueda heurística
  - Planificación y programación

- **Apoyo a la decisión**

Jerarquía de palabras clave de Inria para el ámbito de la IA

No proporcionamos definiciones de la IA y de los subdominios: hay abundante literatura al respecto. También se pueden encontrar buenas definiciones en Wikipedia, por ejemplo:

https://en.wikipedia.org/wiki/Artificial_intelligence

https://en.wikipedia.org/wiki/Machine_learning

https://en.wikipedia.org/wiki/Robotics

https://en.wikipedia.org/wiki/Natural_language_processing

https://en.wikipedia.org/wiki/Semantic_Web

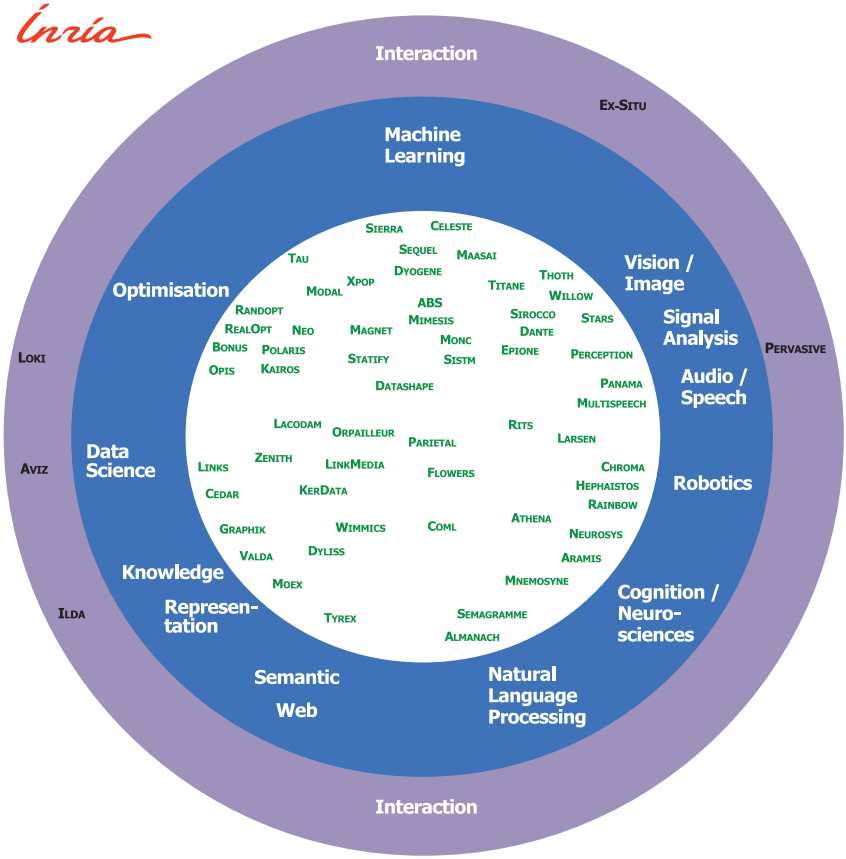
https://en.wikipedia.org/wiki/Knowledge_representation_and_reasoning

etc.

A continuación, las contribuciones de Inria se identificarán por equipos de proyecto.

Los equipos de proyecto de Inria son autónomos, interdisciplinarios y basados en la colaboración y están formados por una media de 15 a 20 integrantes. Los equipos de proyecto se crean sobre la base de una hoja de ruta para la investigación y la innovación y se evalúan al cabo de cuatro años, como parte de una evaluación nacional de todos los equipos de proyecto científicamente similares. Cada equipo es una unidad ágil para llevar a cabo investigaciones de alto riesgo y un caldo de cultivo para las iniciativas empresariales. Dado que las nuevas ideas y las innovaciones disruptivas suelen surgir en la confluencia de varias disciplinas, el modelo de equipo-proyecto promueve el diálogo entre una amplia variedad de métodos, habilidades y áreas temáticas. Como el impulso colectivo es un factor determinante, el 80% de los equipos de investigación de Inria se alían con las principales universidades de investigación y otras organizaciones (CNRS, Inserm, INRAE, etc.) La duración máxima de un equipo-proyecto es de doce años.

Los nombres de los equipos de proyecto se escribirán en MAYÚSCULAS, para distinguirlos de otros sustantivos.



Tras una subsección inicial que trata de los retos genéricos, se presentan retos más específicos, empezando por el aprendizaje automático y siguiendo por las categorías del diagrama circular anterior. El diagrama tiene tres partes: en el interior, los equipos de proyectos; en el anillo interior, las subcategorías de IA; en el anillo exterior, los equipos de interacción hombre-computadora con IA. Cada sección está dedicada a una categoría, y comienza con una copia del diagrama en la que los equipos identificados como pertenecientes a esa categoría están subrayados en azul oscuro y los equipos que tienen una relación más débil con esa categoría están subrayados en azul claro.

5.1 Retos genéricos de la inteligencia artificial

Algunos ejemplos de los principales retos genéricos de la IA identificados por Inria son los siguientes:

Adaptación conjunta y confiable entre los seres humanos y los sistemas basados en la IA. Los datos están presentes en todos los entornos personales y profesionales. Los tratamientos y decisiones basados en algoritmos sobre estos datos se están difundiendo en todos los ámbitos de actividad, con enormes repercusiones en nuestra economía y organización social. La transparencia y la ética de estos sistemas algorítmicos, en particular los sistemas basados en la IA capaces de tomar decisiones críticas, se han convertido en cualidades cada vez más importantes para la confianza y la apropiación de los servicios digitales. Por lo tanto, el desarrollo de métodos de gestión y análisis de datos transparentes y responsables, orientados a los seres humanos, constituye una prioridad muy desafiante.

i) Ciencia de datos para todos. A medida que el volumen y la variedad de los datos disponibles van creciendo, la necesidad de darles sentido es cada vez más acuciante. La ciencia de datos, que engloba diversas tareas como la predicción y el descubrimiento de conocimientos, pretende dar respuesta a esta necesidad y suscita un gran interés. Sin embargo, la ejecución de estas tareas exige un gran esfuerzo por parte de los expertos humanos. Por ello, el diseño de métodos de Ciencia de Datos que reduzcan en gran medida tanto la cantidad como la dificultad del trabajo de los expertos humanos constituye un gran reto para los próximos años.

ii) Interacción adaptativa permanente con los humanos. Los sistemas digitales y robóticos interactivos tienen un gran potencial para ayudar a las personas en tareas y entornos cotidianos, con muchas e importantes aplicaciones sociales: cobots que colaboran con los humanos en las fábricas; vehículos con grandes grados de autonomía; robots y sistemas de realidad virtual que ayudan en la educación o a las personas mayores. En todas estas aplicaciones, los sistemas digitales y robóticos interactivos son herramientas que interconectan el mundo real (donde los humanos experimentan interacciones físicas y sociales) con el espacio digital (algoritmos, repositorios de información y mundos virtuales). A veces, estos sistemas son también una interfaz entre humanos, por ejemplo, cuando constituyen herramientas de mediación entre alumnos y profesores en las escuelas, o entre grupos de personas que colaboran e interactúan en una tarea. Su dimensión física y tangible es a menudo esencial tanto para la función para la que están concebidos (que implica una acción física) como para su adecuada percepción y comprensión por parte de los usuarios.

iii) Vehículos autónomos conectados. El vehículo autónomo conectado (VAC) está surgiendo rápidamente como una solución parcial al reto social de la movilidad sostenible. El CAV no debe considerarse de forma aislada, sino como un eslabón esencial de los sistemas inteligentes de transporte (SIT) cuyos beneficios son muchos: mejora de la seguridad y la eficiencia del transporte por carretera, mejora del acceso a la movilidad y preservación del medio ambiente mediante la reducción de las emisiones de gases de efecto invernadero. El objetivo de Inria es contribuir al diseño de arquitecturas de control avanzadas que garanticen una navegación segura de los CAV mediante la integración de componentes de percepción, planificación, control, supervisión y hardware y software seguros. La validación y verificación de los CAV mediante la creación de prototipos avanzados y la implementación in situ se llevará a cabo en cooperación con los socios industriales competentes.

Además de los retos mencionados, se espera que las propiedades para los sistemas de IA tratados a continuación den lugar a nuevas actividades de investigación más allá de las actuales: algunas son extremadamente complejas y no pueden ser abordadas a corto plazo, pero merecen nuestra atención.

Apertura a otras disciplinas

Una IA suele estar integrada en un sistema más amplio compuesto por muchos elementos. Por lo tanto, la apertura significa que los científicos y desarrolladores de IA tendrán que colaborar con especialistas en otras disciplinas de la informática (por ejemplo, modelización, verificación y validación, redes, visualización, interacción persona-computadora, etc.) que conformen el sistema más amplio, así como con científicos no informáticos que contribuyen a la IA, como psicólogos, biólogos (por ejemplo, biomimética), matemáticos, etc. El segundo aspecto a tener en cuenta es el impacto de los sistemas de IA en varias facetas de nuestra vida, economía y sociedad y, por tanto, la necesidad de colaborar con especialistas de otros campos. Sería demasiado largo mencionarlos a todos, pero podemos poner como ejemplo a economistas, ecologistas, biólogos y abogados.

Aumentar la escala... ¡y reducirla!

Los sistemas de IA deben ser capaces de manejar grandes cantidades de datos y de situaciones. Hemos visto algoritmos de aprendizaje profundo que absorben millones de puntos de datos (señales, imágenes, vídeos, etc.) y sistemas de razonamiento a gran escala, como Watson de IBM, que utilizan conocimientos enciclopédicos; sin embargo, la cuestión general de mayor escalado para las múltiples

V (variedad, volumen, velocidad, vocabularios, etc.) sigue vigente.

Trabajar con microdatos es un reto para varias aplicaciones que no se benefician de las grandes cantidades de casos existentes. Los sistemas integrados, con sus restricciones específicas (recursos limitados, tiempo real, etc.), también plantean nuevos retos. Esto es especialmente relevante para varias industrias y exige el desarrollo de nuevos mecanismos de aprendizaje automático, ya sea ampliando las técnicas de aprendizaje (profundo) (por ejemplo, el aprendizaje por transferencia o el aprendizaje con muy pocos datos (few-shot learning), o considerando enfoques completamente diferentes.

Multitarea

Muchos sistemas de IA son buenos en una cosa, pero muestran poca competencia fuera de su ámbito de interés; pero los sistemas de la vida real, como los robots, deben ser capaces de realizar varias acciones en paralelo, como memorizar hechos, aprender nuevos conceptos, actuar en el mundo real e interactuar con los humanos. Pero esto no es tan sencillo. La diversidad de canales a través de los cuales percibimos nuestro entorno, el razonamiento que realizamos, las tareas que llevamos a cabo, es varios órdenes de magnitud mayor. Incluso si introducimos todos los datos del mundo en la mayor computadora imaginable, estaremos muy lejos de las capacidades de nuestro cerebro. Para mejorar, tendremos que hacer que habilidades especializadas cooperen en subproblemas: es el conjunto de estos subsistemas el que podrá resolver problemas complejos. El futuro de la IA descentralizada y de los sistemas multiagentes promete ser brillante.

Validación y certificación

Un componente obligatorio en los sistemas de misión crítica, la certificación de los sistemas de IA, o su validación por los medios adecuados, es un verdadero reto, especialmente si estos sistemas cumplen las expectativas mencionadas anteriormente (adaptación, multitarea, usuario en el bucle): la verificación, la validación y la certificación de los sistemas clásicos (es decir, los que no son de IA) ya son tareas difíciles, aunque ya existan tecnologías aprovechables, algunas de las cuales están siendo desarrolladas por los equipos de proyectos de Inria-, pero la aplicación de estas herramientas a sistemas complejos de IA es una tarea ingente que hay que abordar si queremos poder utilizar estos sistemas en entornos como aviones, centrales nucleares, hospitales, etc.

Además, mientras que la validación requiere comparar un sistema de IA con sus especificaciones, la certificación requiere la presencia de normas y estándares a

los que el sistema se enfrentará. Varias organizaciones, entre ellas la ISO, trabajan ya en la elaboración de normas para la inteligencia artificial, pero se trata de una misión a largo plazo que no ha hecho más que empezar.

Confianza, justicia, transparencia y responsabilidad

Como se ha visto en el capítulo 3, las cuestiones éticas ocupan ahora un lugar central en los debates sobre la IA y son aún más fuertes en el caso del aprendizaje automático. La confianza puede alcanzarse a través de una combinación de muchos factores, entre ellos la solidez demostrada de los modelos, su capacidad de explicación o su interpretabilidad/auditabilidad por parte de los usuarios humanos, la provisión de intervalos de confianza para los resultados. Estos puntos son clave para lograr una mayor aceptación del uso de la IA en aplicaciones críticas como la medicina, el transporte, las finanzas o la defensa. Otra cuestión importante es la equidad, es decir, la construcción de algoritmos y modelos que traten de forma justa a las diferentes categorías de la población. Hay decenas de análisis e informes sobre esta cuestión, pero prácticamente ninguna solución por el momento.

Normas y valores humanos

Dar normas y valores a las IA va mucho más allá de la ciencia y las tecnologías actuales: por ejemplo, ¿debe un robot que va a comprar leche para su dueño detenerse en su camino para ayudar a una persona cuya vida está en peligro? ¿Podría utilizarse una potente tecnología de Inteligencia Artificial por terroristas igualmente artificiales? Como para otras tecnologías, hay numerosas preguntas fundamentales sin respuesta.

Privacidad

La necesidad de privacidad es especialmente relevante para los sistemas de IA que están expuestos a datos personales, como los asistentes/acompañantes inteligentes o los sistemas de extracción de datos. Esta necesidad también es válida para los sistemas que no son de IA, pero la especificidad de la IA es que los nuevos conocimientos se obtendrán de datos privados y posiblemente se harán públicos si no se limitan por medios técnicos. Algunos sistemas de IA nos conocen mejor que nosotros mismos.

5.2 Aprendizaje automático

A pesar de que el aprendizaje automático (ML por sus siglas en inglés) es la tecnología por la que la Inteligencia Artificial alcanzó nuevos niveles de rendimiento y encontró aplicaciones en casi todos los sectores de la actividad humana, siguen existiendo varios desafíos, desde la investigación básica hasta cuestiones sociales, incluyendo la eficiencia del hardware, la hibridación con otros paradigmas, etc.

Esta sección comienza con algunos retos genéricos en el ámbito de la inteligencia artificial: cuestiones éticas y de confianza –incluida la resistencia a los ataques de adversarios–; rendimiento y consumo de energía; modelos híbridos; transición a la causalidad en lugar de a las correlaciones; comprensión del sentido común; aprendizaje continuo; aprendizaje con restricciones. A continuación se incluyen subsecciones sobre aspectos más específicos, como los fundamentos y la teoría del ML, el ML y los datos heterogéneos, el ML para las ciencias de la vida, con la presentación de los equipos de proyecto Inria.

Resistir a los ataques adversarios

En los últimos años se ha demostrado que los modelos de ML son muy débiles con respecto a los ataques de adversarios, es decir, es bastante fácil engañar a un modelo de aprendizaje profundo modificando ligeramente su señal de entrada y obteniendo así clasificaciones o predicciones erróneas. Resistir tales ataques adversarios es obligatorio para los sistemas que se utilizarán en la vida real, pero una vez más, aún deben desarrollarse soluciones genéricas.

Rendimiento y consumo de energía

Como se muestra en el último informe de IA[19] y en una serie de artículos recientes, la demanda de recursos de cómputo para entrenamiento de aprendizaje automático ha crecido exponencialmente desde 2010, duplicándose cada 3,5 meses - esto significa que se ha multiplicado por mil en tres años, por un millón en seis años. Esto se debe al tamaño de los datos utilizados, a la sofisticación de los modelos de aprendizaje profundo con miles de millones de parámetros como mínimo, y a la aplicación de algoritmos de búsqueda automática de arquitectura que básicamente consisten en ejecutar miles de variaciones de los modelos sobre los mismos datos. El trabajo de Strubell y otros[20] muestra que la energía utilizada

19. Raymond Perrault et al., The AI Index 2019 Annual Report, AI Index Steering Committee, Human-Centered AI Institute, Stanford University, Stanford, CA, diciembre de 2019.
20. Energy and Policy Considerations for Deep Learning in NLP; Strubell, Ganesh, McCallum; College of Information and Computer Sciences, University of Massachussets Amherst, junio de 2019, arXiv:1906.02243v1.

para entrenar un modelo de transformer para el procesamiento del lenguaje natural con búsqueda de arquitectura es cinco veces mayor que el combustible utilizado por un coche de pasajeros medio durante su vida útil. Evidentemente, esto no es sostenible: ahora se oyen voces que exigen revisar la forma en que las máquinas aprenden para ahorrar recursos computacionales y energía. Una idea es la de redes neuronales con conexiones parsimoniosas bajo algoritmos robustos y matemáticamente bien entendidos, lo que lleva a un compromiso entre rendimiento y frugalidad. También se trata de garantizar la robustez de los enfoques, así como la interpretabilidad y explicabilidad de las redes aprendidas.

Modelos híbridos, representaciones simbólicas versus continuas

La hibridación consiste de unir diferentes enfoques de modelización en sinergia: los enfoques más comunes son las representaciones continuas utilizadas para el aprendizaje profundo, los enfoques simbólicos de la antigua comunidad de IA (sistemas expertos y basados en el conocimiento), y los modelos numéricos desarrollados para la simulación y optimización de sistemas complejos. Los partidarios de esta hibridación afirman que dicha combinación, aunque no es fácil de implementar, es mutuamente beneficiosa. Por ejemplo, las representaciones continuas son diferenciables y permiten a los algoritmos de aprendizaje automático abordar funciones complejas, mientras que las representaciones simbólicas se utilizan para aprender reglas y modelos simbólicos. Una característica deseada es integrar el razonamiento en la representación continua, es decir, encontrar formas de hacer inferencias sobre los datos numéricos; por otro lado, para beneficiarse de la potencia del aprendizaje profundo, definir representaciones continuas de los datos simbólicos puede ser muy útil, como se ha hecho, por ejemplo, para el texto con representaciones word2vec y text2vec.

Pasando a la causalidad

Los algoritmos de aprendizaje más utilizados correlacionan los datos de entrada y salida, por ejemplo, entre los píxeles de una imagen y un indicador de una categoría como “gato”, “perro”, etc. Esto funciona muy bien en muchos casos, pero ignora la noción de causalidad, que es esencial para construir sistemas prescriptivos. La causalidad es una herramienta formidable para fabricar este tipo de instrumentos, indispensables para supervisar y controlar sistemas críticos como una central nuclear, el estado de salud de un ser vivo o un avión. Insertar la noción de causalidad en los algoritmos de aprendizaje automático es un reto fundamental; puede hacerse integrando conocimientos a priori (modelos numéricos, lógicos, simbólicos, etc.) o descubriendo la causalidad en los datos.

Comprensión del sentido común

Aunque el rendimiento de los sistemas de aprendizaje automático en términos de tasas de error en varios problemas es bastante notable, se dice que estos modelos no desarrollan una comprensión profunda del mundo, a diferencia de los humanos. La búsqueda de la comprensión del sentido común es larga y tediosa, y comenzó con enfoques simbólicos en los años ochenta y continuó con enfoques mixtos como IBM Watson, el proyecto de robot TODAI[21] (hacer que un robot apruebe un examen para entrar en la Universidad de Tokio), el proyecto Aristo de AllenAI[22] (construir sistemas que demuestren una comprensión profunda del mundo, integrando tecnologías de lectura, aprendizaje, razonamiento y explicación) y, más recientemente, el proyecto Debater de IBM[23], un sistema capaz de intercambiar argumentos sobre cualquier tema con los mejores oradores humanos. Un sistema como Meena[24] de Google (un agente conversacional que puede charlar sobre cualquier cosa) puede crear una cierta expectativa cuando lo vemos conversar, pero la comprensión profunda de sus conversaciones es otra cuestión diferente.

Aprendizaje continuo e interminable (de por vida)

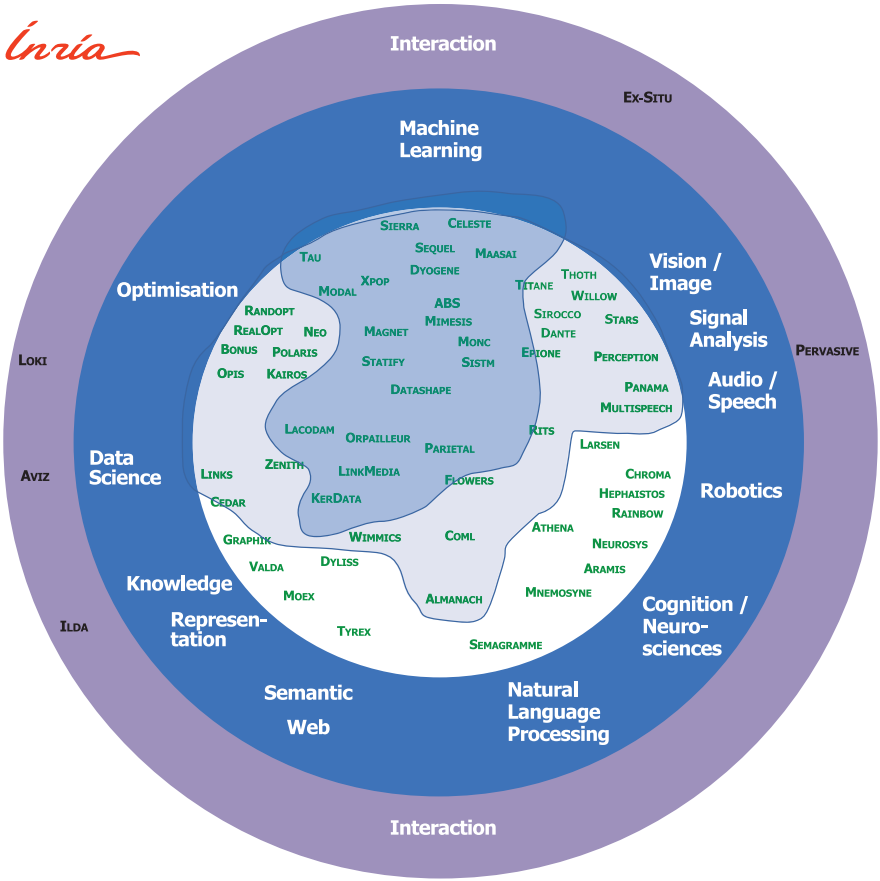
Se espera que algunos sistemas de IA sean resilientes, es decir, que puedan funcionar las 24 horas del día sin interrupciones. Se han producido interesantes avances en los sistemas de aprendizaje permanente que permiten acumular continuamente nuevos conocimientos mientras operan. El desafío aquí reside en la capacidad de los sistemas de IA para operar en línea en tiempo real y ser capaces de desafiar las creencias previas de forma autónoma. Estos sistemas utilizan algún tipo de arranque inicial y son autosuficientes (bootstrapping): los conocimientos elementales aprendidos en las primeras etapas de funcionamiento se utilizan para dirigir futuras tareas de aprendizaje, como en el sistema NELL/Read the Web (aprendizaje de lenguaje sin fin) desarrollado por Tom Mitchell en la Universidad de Carnegie-Mellon[25].

Aprendizaje con restricciones

La privacidad es, sin duda, la restricción más importante que hay que tener en

21. https://21robot.org/index-e.html
22. https://allenai.org/aristo
23. https://www.research.ibm.com/artificial-intelligence/project-debater/
24. https://ai.googleblog.com/2020/01/towards-conversational-agent-that-can.html
25. http://rtw.ml.cmu.edu/rtw/

cuenta. En el campo del aprendizaje automático se ha asumido recientemente la necesidad de mantener la privacidad mientras se aprende de los registros sobre los individuos; los investigadores están desarrollando una teoría del aprendizaje automático respetuosa con la privacidad. En Inria, varios equipos trabajan sobre la privacidad: especialmente ORPAILLEUR en el aprendizaje automático, pero también equipos de otros ámbitos como PRIVATICS (algoritmos de la privacidad) y SMIS (privacidad en las bases de datos). En términos más generales, el aprendizaje automático debería tener en cuenta a otras restricciones externas, como la descentralización de los datos o las limitaciones energéticas, como se ha mencionado anteriormente. Por lo tanto, es necesario investigar el problema general del aprendizaje automático con restricciones externas.



5.2.1 Aprendizaje automático fundamental y modelos matemáticos

El aprendizaje automático plantea numerosas cuestiones fundamentales, como la vinculación de la teoría con la experimentación, la generalización, la capacidad de explicar el resultado del algoritmo, pasar al aprendizaje no supervisado o poco supervisado, etc. También hay cuestiones relacionadas con las infraestructuras informáticas y, como se ha visto en el apartado anterior, con el uso de los recursos de cómputo. Varios equipos de Inria trabajan en el aprendizaje automático fundamental, desarrollando nuevos conocimientos matemáticos y aplicándolos a casos de uso del mundo real.

Teoría matemática

Los algoritmos de aprendizaje se basan en sofisticadas matemáticas, lo que dificulta su comprensión, uso y explicación. Uno de los retos es mejorar los fundamentos teóricos de nuestros modelos, que a menudo se ven externamente como cajas negras algorítmicas difíciles de interpretar. Conseguir que la teoría y la práctica se compenetren al máximo es un reto constante, y uno que cada vez es más importante dado el número de investigadores e ingenieros aplicados que trabajan en IA/aprendizaje automático: Los métodos "actuales" en la práctica se alejan constantemente de lo que la teoría puede justificar o explicar.

Generalización

Un reto central del aprendizaje automático es el de la generalización: cómo una máquina puede predecir/controlar un sistema más allá de los datos que ha visto durante el entrenamiento, especialmente más allá de la distribución de los datos observados durante el entrenamiento. Además, la generalización ayudará a pasar de sistemas que pueden resolver una tarea a sistemas polivalentes que pueden aplicar sus capacidades en diferentes contextos. Esto también puede ser por transferencia (de una tarea a otra) o por adaptación.

Explicabilidad

Uno de los factores de confianza en los sistemas artificiales, la explicabilidad es necesaria para los sistemas que realizan predicciones y decisiones críticas, cuando no existen otras garantías como la verificación formal, la certificación o la adhesión

a normas y estándares[26]. La búsqueda de la explicabilidad de los sistemas de IA lleva ya tiempo en marcha; fue desencadenada por el programa XAI (eXplainable AI) de DARPA[27], lanzado en 2017. Hay muchos intentos de ofrecer explicaciones (por ejemplo, resaltando determinadas áreas en las imágenes, realizando análisis de sensibilidad en los datos de entrada, transformando parámetros numéricos en símbolos o reglas condicionales tipo if-then), pero ninguno es totalmente satisfactorio.

Coherencia de los resultados o salidas de los algoritmos

Se trata de un requisito previo para el desarrollo de los marcos legales necesarios para las pruebas a gran escala y el despliegue de los vehículos autónomos (AV por sus siglas en inglés) en las redes de carreteras y ciudades reales. El problema de la reproducibilidad estadística: poder asignar un nivel de significación (por ejemplo, un valor p) a las conclusiones extraídas de un algoritmo de aprendizaje automático. Esta información parece indispensable para fundamentar el proceso de toma de decisiones basado en estas conclusiones.

Programación diferenciable

Más allá de la disponibilidad de datos y de los potentes computadoras que explican la mayoría de los recientes avances en el aprendizaje profundo, hay una tercera razón que es a la vez científica y tecnológica: hasta 2010, los investigadores en aprendizaje automático derivaron las fórmulas analíticas para calcular los gradientes con el método de retro-propagación, backprogramation en su acepción inglesa. Entonces redescubrieron la diferenciación automática, que existía en otras comunidades pero que aún no había entrado en el campo de la IA. Esto abrió la posibilidad de experimentar con arquitecturas complejas como los Transformers/BERTs que revolucionaron el procesamiento del lenguaje natural. Hoy podríamos sustituir el término "aprendizaje profundo" por "programación diferenciable", que es a la vez más científico y más genérico.

26. Algunos especialistas en Aprendizaje Profundo afirman que la gente confía en sus médicos sin explicaciones, lo cual es cierto. Pero los médicos siguen un largo periodo de formación materializado en un diploma que certifica sus capacidades.
27. https://www.darpa.mil/program/explainable-artificial-intelligence

CELESTE

Estadísticas matemáticas y aprendizaje

La comunidad estadística tiene una larga experiencia en la forma de inferir conocimientos a partir de los datos, basada en sólidos fundamentos matemáticos. El campo más reciente del aprendizaje automático también ha hecho importantes progresos al combinar la estadística y la optimización, con un punto de vista novedoso que tiene su origen en aplicaciones donde la predicción es más importante que la construcción de modelos.

El equipo-proyecto Celeste se sitúa en la intersección entre la estadística y el aprendizaje automático. Son estadísticos en un departamento de matemáticas, con una sólida formación matemática detrás, interesados en las interacciones entre teoría, algoritmos y aplicaciones. De hecho, las aplicaciones son la fuente de muchos de nuestros interesantes problemas teóricos, mientras que la teoría que desarrollamos desempeña un papel clave en (i) la comprensión de cómo y por qué funcionan los algoritmos de aprendizaje estadístico de éxito – y, por lo tanto, mejorarlos - y (ii) la construcción de nuevos algoritmos sobre bases matemáticas basadas en la estadística.

El objetivo de Celeste es analizar los algoritmos de aprendizaje estadístico - especialmente los más utilizados en la práctica - con nuestro punto de vista de la estadística matemática, y desarrollar nuevos algoritmos de aprendizaje basados en nuestros conocimientos de estadística matemática.

Los objetivos teóricos y metodológicos de Celeste corresponden a cuatro grandes retos del aprendizaje automático en los que la estadística matemática tiene un papel fundamental:

■ En primer lugar, cualquier proceso de aprendizaje automático depende de unos hiperparámetros que hay que elegir, y hay muchos procedimientos disponibles para cualquier problema de aprendizaje: ambos son un problema de selección de estimadores.

■ En segundo lugar, con datos de alta dimensión y/o de gran tamaño, la complejidad computacional de los algoritmos debe tenerse en cuenta

de forma diferente, lo que lleva a posibles disyuntivas entre la precisión y la complejidad estadística, tanto para los propios procedimientos de aprendizaje automático como para los procedimientos de selección de estimadores.

■ En tercer lugar, los datos reales suelen estar parcialmente corruptos, por lo que es necesario proporcionar procesos de aprendizaje (y de selección de estimadores) que sean robustos frente a los valores atípicos y las colas pesadas, y que al mismo tiempo sean capaces de manejar grandes conjuntos de datos.

■ En cuarto lugar, la ciencia se enfrenta actualmente a una crisis de reproducibilidad, por lo que es necesario proporcionar herramientas estadísticas inferenciales (valores p, márgenes de confianza) para evaluar la significancia de los resultados de cualquier algoritmo de aprendizaje (incluyendo el ajuste de sus hiperparámetros), de una manera computacionalmente eficiente.

TAU

Abordar lo sub-especificado

Aprovechando la experiencia en aprendizaje automático (ML) y optimización del equipo de TaO, el proyecto TaU aborda algunos de **los retos sub-especificados que hay detrás de la nueva ola de la inteligencia artificial.**

1. Una IA de confianza

El temor a los efectos no deseados de la IA y el aprendizaje automático obedece a tres razones (i) cuanto más inteligente es el sistema, más complejo es y más difícil es corregir sus errores (problema de certificación); ii) si el sistema aprende a partir de datos que reflejan los sesgos del mundo (prejuicios, desigualdades), los modelos aprendidos tenderán a perpetuar estos sesgos (problemas de equidad); iii) la IA y el aprendizaje tienden a aprender a partir de modelos predictivos (si se dan las condiciones, aparecen los efectos); y los responsables de la toma de decisiones tienden a utilizar estos modelos de forma prescriptiva (para producir esos efectos, hay que tratar de satisfacer esas condiciones), lo que puede ser ineficaz o

incluso catastrófico (problemas de causalidad). Certificación de modelos. Un posible enfoque para certificar las redes neuronales se basa en pruebas formales. El principal obstáculo en este caso es la etapa de percepción, para la que no existe una especificación formal o una descripción manejable del conjunto de escenarios posibles. Una posibilidad es considerar que el conjunto de escenarios/percepciones es captado por un simulador, lo que permite limitarse a un problema muy simplificado, pero bien fundamentado.

Sesgo y equidad. En las ciencias sociales y las humanidades (por ejemplo, los vínculos entre la salud de una empresa y el bienestar de sus empleados, la recomendación de ofertas de trabajo, los vínculos entre la alimentación y la salud) se ofrecen datos sesgados. Por ejemplo, los datos sobre el comportamiento se recogen a menudo con fines de marketing, por lo que pueden tender a sobre representar una u otra categoría. Estos sesgos deben identificarse y ajustarse para obtener modelos precisos.

Causalidad. Los modelos predictivos pueden basarse en correlaciones (la presencia de libros en casa está correlacionada con las buenas notas de los niños en la escuela). Sin embargo, estos modelos no permiten incidir para conseguir los efectos deseados (por ejemplo, es inútil enviar libros para mejorar las notas de los niños): sólo los modelos causales permiten realizar intervenciones fundadas. La búsqueda de modelos causales abre grandes perspectivas (por ejemplo, el poder del modelo causal para influir en el resultado). La búsqueda de modelos causales abre perspectivas importantes (poder modelar lo que habría pasado si se hubiera hecho de otra manera, es decir, la modelización contrafactual) para la "IA del bien".

2. Aprendizaje automático y computación científica

Un reto clave es combinar el aprendizaje automático y la IA con el conocimiento del dominio. En el campo de la modelización matemática y el análisis numérico, en particular, existen amplios conocimientos de descripción, simulación y diseño en forma de ecuaciones en derivadas parciales. El acoplamiento entre las redes neuronales y los modelos numéricos es una dirección de investigación estratégica, con primeros resultados en cuanto a i) la complejidad de los fenómenos subyacentes (mecánica de fluidos multifásica en 3D, materiales hiperelásticos heterogéneos, etc.); ii) el escalado (simulación en tiempo real); iii) el control fino/adaptativo de

modelos y procesos, por ejemplo, el control de inestabilidades numéricas o la identificación de invariantes físicas.

3. Una IA sostenible: aprender a aprender

El talón de Aquiles del aprendizaje automático, aparte de algunos ámbitos como el procesamiento de imágenes, sigue siendo la dificultad de ajustar los modelos (típicamente para las redes neuronales, pero en términos generales). La calidad de los modelos depende del ajuste automático de toda la cadena de aprendizaje, del pre-procesamiento de datos a parámetros estructurales del propio aprendizaje, de la elección de la arquitectura para las redes profundas, de los algoritmos para el aprendizaje estadístico clásico y de los hiperparámetros de todos los componentes de la cadena de procesamiento.

Los enfoques propuestos van desde los métodos derivados de la teoría de la información y la física estadística hasta los propios métodos de aprendizaje. En el primer caso, dado el gran tamaño de las redes consideradas, se pueden utilizar métodos de física estadística (por ejemplo, campo medio, invariancia de escala) para ajustar los hiperparámetros de los modelos y para definir las áreas del problema en las que se pueden encontrar soluciones. En el segundo caso, se trata de modelizar, a partir del comportamiento empírico, qué algoritmos se comportan bien con qué datos.

Una dificultad relacionada con esto es la cantidad astronómica de datos que se necesitan para aprender los modelos más eficientes del momento, es decir, las redes neuronales profundas. El costo computacional se convierte así en un obstáculo importante para la reproducibilidad de los resultados científicos.

Aprendizaje semi-supervisado y no supervisado

La mayoría de los resultados notables obtenidos con el ML se basan en el aprendizaje supervisado, es decir, el aprendizaje a partir de ejemplos en los que la salida esperada se da con los datos de entrada. Esto implica el etiquetado previo de los datos con la correspondiente salida esperada y puede ser bastante exigente para los datos a gran escala. El Mechanical Turk de Amazon es un ejemplo de cómo las empresas movilizan recursos humanos para registrar datos (lo que plantea muchos problemas sociales). Aunque el aprendizaje supervisado ofrece

sin duda un rendimiento excelente, el coste del etiquetado acaba siendo insostenible ya que aumenta el tamaño de los conjuntos de datos. Tampoco es práctico incluir todas las condiciones de funcionamiento en un único conjunto de datos. El aprendizaje semi-supervisado o no supervisado es necesario para garantizar la escalabilidad de los algoritmos en el mundo real, donde eventualmente se enfrentarán a situaciones no vistas en el conjunto de entrenamiento. El Santo Grial de la inteligencia artificial general está lejos de nuestros conocimientos actuales, pero las prometedoras técnicas de aprendizaje por transferencia permiten extender el entrenamiento realizado de forma supervisada a nuevos conjuntos de datos no etiquetados, por ejemplo con la adaptación de dominios.

Arquitecturas de Cómputo

Los sistemas modernos de aprendizaje automático necesitan recursos de cómputo de alto rendimiento y un almacenamiento de datos para poder escalar con el tamaño de los datos y las dimensiones del problema; los algoritmos se ejecutarán en unidades de procesamiento gráfico (GPU) y otras arquitecturas potentes como las unidades de procesamiento tensorial (TPU), las unidades de procesamiento neuronal (NPU), las unidades de procesamiento de inteligencia (IPU), etc.; los datos y los procesos deben distribuirse entre muchos procesadores. La investigación futura debería centrarse en cómo pueden mejorarse los algoritmos de ML y las formulaciones de los problemas para aprovechar al máximo estas arquitecturas de cómputo, atendiendo también a cuestiones de sostenibilidad (véase más arriba).

MAASAI

Modelos y algoritmos para la inteligencia artificial

Maasai es un equipo-proyecto de investigación de Inria Sophia-Antipolis que trabaja en los modelos y algoritmos de la Inteligencia Artificial. Se trata de un equipo de investigación conjunto con los laboratorios LJAD (Matemáticas, UMR 7351) e I3S (Informática, UMR 7271) de la Université Côte d'Azur. El equipo está formado por matemáticos e informáticos con el fin de proponer metodologías de aprendizaje innovadoras, que aborden problemas del mundo real, y que sean a la vez teóricamente sólidas, escalables y asequibles.

La inteligencia artificial se ha convertido en un elemento clave en la mayoría

de los campos científicos y ya forma parte de la vida de todos gracias a la revolución digital. Los métodos estadísticos, de aprendizaje automático y profundo intervienen en la mayoría de las aplicaciones científicas en las que hay que tomar una decisión, como el diagnóstico médico, los vehículos autónomos o el análisis de textos. Los recientes y muy publicitados resultados de la inteligencia artificial no deben ocultar los subsistentes y nuevos problemas que plantean los datos modernos. En efecto, a pesar de las recientes mejoras debidas al aprendizaje profundo, la naturaleza de los datos modernos ha planteado problemas específicos. Por ejemplo, el aprendizaje con datos de alta dimensión, atípicos (redes, funciones, etc.), dinámicos o heterogéneos sigue siendo difícil por razones teóricas y algorítmicas. El reciente surgimiento del aprendizaje profundo también ha abierto nuevas cuestiones como: ¿Cómo aprender en un contexto no supervisado o semi-supervisado con arquitecturas profundas? ¿Cómo diseñar una arquitectura profunda para una situación determinada? ¿Cómo aprender con datos cambiantes y corruptos?

Para abordar estas cuestiones, el equipo de Maasai se centra en temas como el aprendizaje no supervisado, la teoría del aprendizaje profundo, el aprendizaje adaptativo y robusto, y el aprendizaje con datos de alta dimensión o heterogéneos. El equipo Maasai lleva a cabo una investigación que vincula problemas prácticos que pueden provenir de la industria o de otros campos científicos, con los aspectos teóricos de las Matemáticas y la Informática. Con este espíritu, el equipo-proyecto Maasai está totalmente integrado en el eje "Core elements of AI";"Elementos básicos de la IA", del Institut 3IA Côte d'Azur. Cabe destacar que el equipo acoge dos cátedras 3IA del Institut 3IA Côte d'Azur.

SIERRA

Aprendizaje automático estadístico y parsimonia

SIERRA aborda principalmente problemas de aprendizaje automático, con el objetivo principal de establecer un vínculo entre la teoría y los algoritmos, y entre los algoritmos y las aplicaciones de alto impacto en diversos campos de la ingeniería y la ciencia, en particular la visión por computadora, la bioinformática, el procesamiento de audio, el procesamiento de textos

y la neuroimagenología.

Los logros recientes incluyen trabajos teóricos y algorítmicos para la optimización convexa a gran escala, lo que conduce a algoritmos que realizan pocas pasadas sobre los datos y, al mismo tiempo, logran un rendimiento predictivo óptimo en una amplia variedad de situaciones de aprendizaje supervisado. Los retos para el futuro incluyen el desarrollo de nuevos métodos para el aprendizaje no supervisado, el diseño de algoritmos de aprendizaje para arquitecturas de cómputo paralelas y distribuidas, y la comprensión teórica del aprendizaje profundo.

Desafíos en el aprendizaje por refuerzo

Hacer más efectivo el aprendizaje por refuerzo permitiría acometer tareas realmente relevantes, especialmente las estocásticas y no estacionarias. Para ello, las tendencias actuales son utilizar el aprendizaje por transferencia entre tareas, y la posibilidad de integrar conocimiento previo.

Aprendizaje por transferencia

El aprendizaje por transferencia es útil cuando hay pocos datos disponibles para aprender una tarea. Implica utilizar para una nueva tarea lo que se ha aprendido de otra tarea sobre la que se dispone de más datos. Es una idea bastante antigua (1993), pero sus resultados son modestos porque su aplicación es difícil: implica abstraer lo que el sistema ha aprendido en primer lugar, pero no existe una solución general a este problema (¿qué abstraer, cómo, cómo reutilizar?, etc.). Otro enfoque del aprendizaje por transferencia es el procedimiento conocido como "shaping": aprender una tarea sencilla, y gradualmente hacerla más compleja, hasta llegar a la tarea objetivo. Hay ejemplos de este proceso en la literatura, pero no hay una teoría general.

SCOOL

El equipo-proyecto SCOOL (antes conocido como SEQUEL) trabaja en el campo del aprendizaje automático digital. El objetivo de SCOOL es estudiar los problemas de toma de decisiones secuenciales en condiciones de

incertidumbre, en particular los problemas de bandidos y problemas de aprendizaje por refuerzo.

Las actividades de SCOOL abarcan desde la investigación básica hasta las aplicaciones y la transferencia de tecnología. En cuanto a la investigación básica y formal, SCOOL se centra en la modelización de problemas concretos, el diseño de nuevos algoritmos y el estudio de las propiedades formales de estos algoritmos (convergencia, velocidad, eficiencia, entre otros). En un terreno más algorítmico, participan en los esfuerzos relativos a la mejora de los algoritmos de aprendizaje por refuerzo para la resolución de tareas más amplias y estocásticas. Este tipo de tareas incluye, naturalmente, el problema de la gestión de recursos limitados con el fin de realizar de la mejor manera posible una tarea determinada. SCOOL ha sido muy activo en el ámbito de los sistemas de recomendación en línea. En los últimos años, su trabajo ha dado lugar a aplicaciones en tareas de aprendizaje de diálogo en lenguaje natural y visión por computadora. Actualmente, hacen especial hincapié en la resolución de estos problemas en entornos no estacionarios, es decir, entornos cuya dinámica cambia con el tiempo.

En la actualidad, SCOOL centra sus esfuerzos en aplicaciones en los ámbitos de la salud, la educación y el desarrollo sostenible (gestión de la energía, por un lado, y agricultura, por otro).

DYOGENE

Dinámica de las redes geométricas

El enfoque científico de DYOGENE se centra en la dinámica de las redes geométricas que surgen en las comunicaciones. Las redes geométricas abarcan redes con una definición geométrica de la existencia de conexiones entre los nodos, como los grafos aleatorios y las redes geométricas estocásticas.

- **Aprendizaje no supervisado para datos estructurados en grafos**

En muchos escenarios, los datos se representan de forma natural como un grafo, ya sea directamente (por ejemplo, las interacciones entre los agentes en una red social en línea), o después de algún tipo de procesamiento (por

ejemplo, el grafo de vecindad más cercano entre las palabras incrustadas en algún espacio euclidiano). Entre las tareas fundamentales de aprendizaje no supervisado para este tipo de datos gráficos se encuentran el agrupamiento de grafos y el alineamiento de grafos.

DYOGENE desarrolla algoritmos de gran eficacia para llevar a cabo estas tareas, centrándose en escenarios difíciles en los que la cantidad de ruido en los datos es elevada, de modo que los métodos clásicos fallan. En particular, investigan: métodos espectrales, algoritmos de transmisión de mensajes y redes neuronales de grafos.

- **Aprendizaje automático distribuido**

El aprendizaje automático moderno requiere procesar conjuntos de datos distribuidos en varias máquinas, ya sea porque no caben en una sola o por restricciones de privacidad. DYOGENE desarrolla algoritmos novedosos para estos escenarios de aprendizaje disperso que utilizan eficientemente los recursos de comunicación entre las ubicaciones de datos, y los recursos de almacenamiento e informática en dichas ubicaciones de datos.

- **Redes de energía**

DYOGENE desarrolla esquemas de control para el funcionamiento eficiente de las redes de energía, que implican, en particular, métodos de aprendizaje por refuerzo y algoritmos de correspondencia en línea.

5.2.2 Datos heterogéneos/complejos y modelos híbridos

Además de los retos generales del ML vistos anteriormente, los retos para los equipos que ponen el énfasis en los datos consisten en aprender de datos heterogéneos, disponibles a través de múltiples canales; considerar la intervención humana en el bucle de aprendizaje; trabajar con datos distribuidos en la red; trabajar con fuentes de conocimiento además de fuentes de datos, integrando modelos y ontologías en el proceso de aprendizaje (ver la sección 5.4); y, por último, obtener un buen rendimiento de aprendizaje con

pocos datos, en los casos en que las fuentes de big data no son habituales.

Datos heterogéneos

Los datos pueden obtenerse de muchas fuentes: de bases de datos diseminadas a través de Internet o de sistemas de información corporativos; de sensores en el Internet de las cosas; de vehículos conectados; de grandes equipos experimentales, por ejemplo, en la ciencia de los materiales o la astrofísica. Trabajar con datos heterogéneos es obligatorio, independientemente de los medios que se utilicen, es decir, usando directamente la heterogeneidad o estableciendo etapas de preprocesamiento para homogeneizar.

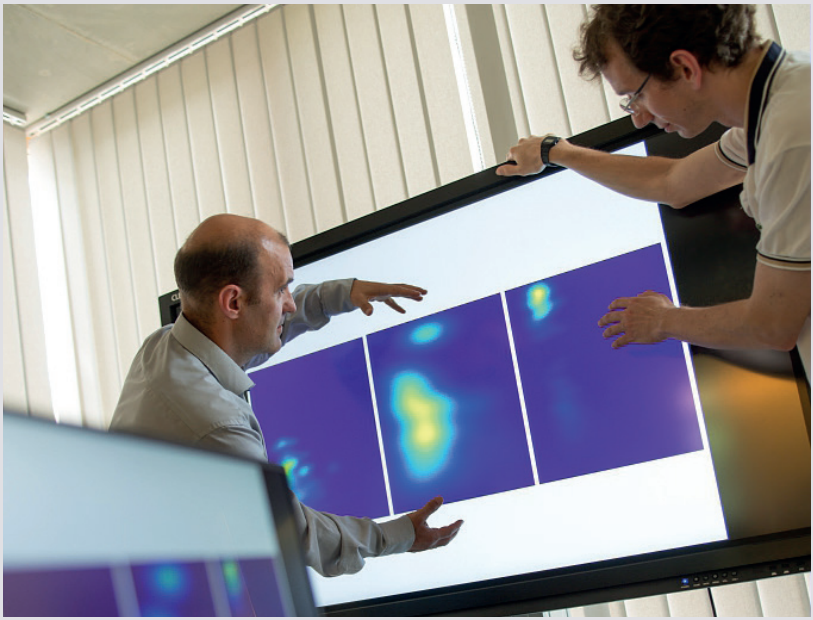
DATASHAPE

Entender la forma de los datos

Los datos complejos modernos, como los datos dependientes del tiempo, las imágenes 3D o los grafos, presentan a menudo una interesante estructura topológica o geométrica. Identificar, extraer y explotar las características topológicas y geométricas o invariantes subyacentes a los datos se ha convertido en un problema de gran importancia para comprender mejor las propiedades relevantes de los sistemas a partir de los cuales se han generado. Partiendo de sólidas bases teóricas y algorítmicas, la inferencia geométrica y la topología computacional han alcanzado progresos importantes en el análisis de datos y el aprendizaje automático. Las nuevas teorías con fundamento matemático dieron lugar al campo del análisis topológico de datos (Topological Data Analysis o, por sus siglas en inglés, TDA), que en la actualidad está suscitando un gran interés tanto del mundo académico como de la industria. Durante los últimos años, el TDA, combinado con otros enfoques de ML e IA, ha sido testigo de muchas contribuciones teóricas exitosas, con la aparición de la teoría de la homología persistente y los enfoques basados en la distancia, importantes desarrollos algorítmicos y de software y aplicaciones exitosas en el mundo real. Estos avances han abierto nuevas vías de investigación teórica, aplicada e industrial en el cruce de TDA, ML e IA.

El equipo de Inria DataShape está llevando a cabo actividades de investigación sobre enfoques topológicos y geométricos en ML e IA con un doble objetivo académico e industrial/social. En primer lugar, basándose en su

sólida experiencia en el Análisis Topológico de Datos, DataShape diseña nuevos métodos y algoritmos topológicos y geométricos matemáticamente bien definidos para el Análisis de Datos y el ML y los pone a disposición de la ciencia de datos y la comunidad de la IA a través de la plataforma de software de última generación GUDHI. En segundo lugar, gracias a las sólidas y duraderas colaboraciones con socios industriales franceses e internacionales, DataShape pretende explotar su experiencia y herramientas para abordar problemas desafiantes con alto impacto social y económico, en particular en la medicina personalizada, el diagnóstico médico asistido por IA o la industria.



Análisis de datos topológicos - © Inria _ Photo C. Morel

MAGNET

Aprendizaje automático en redes de información

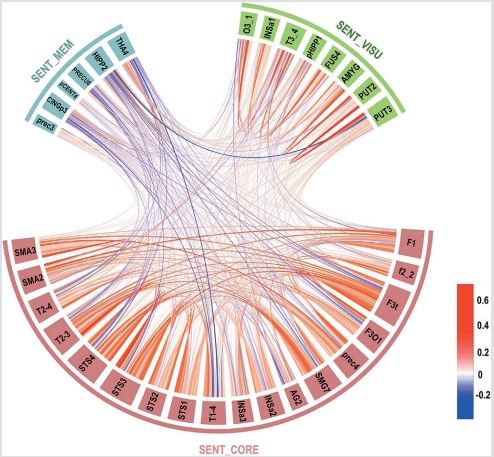
El proyecto Magnet tiene como objetivo diseñar nuevos métodos basados

en el aprendizaje automático orientados a la minería de redes de información. Las redes de información son grandes colecciones de datos y documentos interconectados, como las colecciones de referencias y las redes de blogs, entre otras. Para ello, se definirán nuevos métodos de predicción estructurada para (redes de) textos basados en algoritmos de aprendizaje automático en grafos. Dichos algoritmos incluyen la clasificación de nodos, la predicción de enlaces, el agrupamiento y el modelado probabilístico de grafos. Entre las aplicaciones previstas se encuentran los sistemas de búsqueda, seguimiento y recomendación y, en general, la extracción de información en las redes de información. Los ámbitos de aplicación abarcan las redes sociales de datos culturales y el comercio electrónico, y la informática biomédica. En concreto, los principales objetivos de MAGNET son:

- Aprendizaje de grafos, es decir, construcción, consecución y representación de grafos a partir de datos y de redes (de textos)

- Aprendizaje con grafos, es decir, el desarrollo de técnicas innovadoras para la predicción de enlaces y estructuras en varios niveles de representación (de textos).

Cada punto se estudiará también en contextos en los que se dispone de poca (o ninguna) supervisión. Por lo tanto, a lo largo del proyecto se considerará el aprendizaje semi-supervisado y no supervisado.



Grafo de enlaces de conectividad extrínseca - Labache, L., Joliot, M., Saracco, J. et al. A SENtence Supramodal Areas AtlaS (SENSAAS) based on multiple task-induced activation mapping and graph analysis of intrinsic connectivity in 144 healthy right-handers. Brain Struct Funct 224, 859–882 (2019). Page 870

STATIFY

Modelos estadísticos bayesianos y de valor extremo para datos estructurados y de alta dimensión

El equipo de STATIFY está especializado en la modelización estadística de sistemas con datos de estructura compleja. Ante los nuevos problemas planteados por la ciencia de datos y los métodos de aprendizaje profundo, el objetivo es desarrollar métodos estadísticos matemáticamente bien fundados para proponer modelos que capturen la variabilidad de los sistemas considerados, modelos que sean escalables para procesar datos de alta dimensión y con buenos niveles garantizados de exactitud y precisión. Las aplicaciones a las que se dirigen son principalmente la obtención de imágenes del cerebro (o neuroimagenología), la medicina personalizada, el análisis de riesgos medioambientales y las geociencias. STATIFY es, por tanto, un proyecto científico centrado en la estadística y que desea tener un fuerte impacto metodológico y de aplicación en la ciencia de datos.

STATIFY es la continuación natural del equipo MISTIS. Este nuevo proyecto STATIFY se basa claramente en todas las técnicas desarrolladas en MISTIS, pero consolida o introduce nuevas direcciones de investigación relativas a la modelización bayesiana, los modelos gráficos probabilísticos, los modelos para datos de alta dimensión y, por último, los modelos para la obtención de imágenes cerebrales, proyectos que han estado vinculados a la llegada de dos nuevos miembros permanentes, Julyan Arbel (en septiembre de 2016) y Sophie Achard (en septiembre de 2019).

Este nuevo equipo se centre en el tema "Optimización, aprendizaje y métodos estadísticos" del dominio "Matemáticas aplicadas, cálculo y simulación". Se trata de un proyecto-equipo conjunto entre Inria, Grenoble INP, Université Grenoble Alpes y CNRS, a través de la afiliación del equipo al Laboratorio Jean Kuntzmann, UMR 5224.

El hombre en el bucle de aprendizaje, explicaciones

Los retos se centran en la cooperación sin fisuras de los algoritmos de ML y los usuarios para mejorar el proceso de aprendizaje; para ello, los sistemas de

aprendizaje automático deben ser capaces de presentar sus progresos de forma comprensible para los humanos. Además, el usuario humano debe poder obtener explicaciones del sistema sobre cualquier resultado obtenido. Estas explicaciones se producirían durante la progresión del sistema y podrían estar vinculadas a los datos de entrada o a las representaciones intermedias; también podrían indicar los niveles de confianza según el caso.

LACODAM

Minería de datos colaborativa a gran escala

El objetivo del equipo de Lacodam es facilitar el proceso de dar sentido a (grandes) cantidades de datos. Esto puede servir para obtener conocimientos e ideas para mejorar la toma de decisiones. El equipo estudia principalmente enfoques que proporcionen nuevas herramientas a los científicos de datos, que puedan realizar tareas que no son abordadas por ninguna otra herramienta, o que mejoren el rendimiento en algún área para las tareas existentes (por ejemplo, reducir el tiempo de ejecución, mejorar la exactitud o manejar mejor los datos descompensados).

Una de las principales áreas de investigación del equipo son los métodos novedosos para descubrir patrones dentro de los datos. Estos métodos pueden enmarcarse en los campos de la minería de datos (para el análisis exploratorio de los mismos) o del aprendizaje automático (para tareas supervisadas como la clasificación).

Otro interés clave de la investigación del equipo es el de los métodos de aprendizaje automático interpretables. Hoy en día, hay muchos enfoques de aprendizaje automático que tienen excelentes rendimientos, pero que son muy complejos: sus decisiones no pueden ser explicadas a los usuarios humanos. Una interesante línea de trabajo reciente consiste en combinar el rendimiento en la tarea de aprendizaje automático con la capacidad de justificar las decisiones de forma comprensible. Puede hacerse, por ejemplo, con métodos de interpretabilidad post-hoc, que para una determinada decisión del modelo de aprendizaje automático complejo estimarán su superficie de decisión (compleja) en torno a ese punto.

Esto puede hacerse con un modelo mucho más simple (por ejemplo, un

modelo lineal), que sea comprensible para los humanos.

Detección y caracterización del comportamiento de los usuarios en el contexto del big data
© Inria _ Photo C. Morel

LINKMEDIA

Creación y explotación de enlaces explícitos entre fragmentos multimedia

LINKMEDIA se centra en la interpretación automática de contenidos multimedia profesionales y sociales en todas sus modalidades. En este marco, la inteligencia artificial se basa tanto en el diseño de modelos de contenido como en los algoritmos de aprendizaje asociados para extraer, describir e interpretar los mensajes editados para los humanos. Con el objetivo de realizar análisis multimedia, LINKMEDIA desarrolla algoritmos de aprendizaje automático basados principalmente en modelos estadísticos y neuronales para extraer estructuras, conocimientos, entidades o hechos de documentos y colecciones multimedia. La multimodalidad y la multimodalidad cruzada para enlazar representaciones simbólicas (por ejemplo, palabras o conceptos en un texto) y observaciones continuas (por ejemplo, imágenes continuas o descriptores de señales) es uno de los retos clave para LINKMEDIA, donde el embeber de redes neuronales aparece como

una prometedora dirección de investigación. La detección de bulos en las redes sociales combinando el procesamiento de imágenes y el del lenguaje natural, la creación de hipervínculos en colecciones de vídeos explotando simultáneamente el contenido hablado y visual, y el análisis interactivo de noticias basado en gráficos de proximidad de contenido son algunos de los temas clave en los que trabaja el equipo.

La analítica "User-in-the-loop", en la que la inteligencia artificial está al servicio de un usuario, también es fundamental para el equipo y plantea retos para la interpretación de contenidos multimedia basados en máquinas supervisadas por humanos: el ser humano tiene que entender las decisiones tomadas por las máquinas y evaluar su fiabilidad, dos aspectos que suponen un reto dados los enfoques actuales basados en datos; el conocimiento y el aprendizaje automático están fuertemente entrelazados en este escenario, de ahí la necesidad de mecanismos que permitan a los expertos humanos inyectar conocimiento en los algoritmos de interpretación de datos; los usuarios malintencionados inevitablemente manipularán los datos para sesgar la interpretación automática a su favor, una situación que el actual aprendizaje automático contra adversarios tiene dificultades para manejar; por último, pero no menos importante, en lugar de medidas objetivas sobre los datos anotados, la evaluación de los algoritmos se está orientando hacia paradigmas de diseño centrados en el usuario que son difíciles de convertir en funciones objetivas para optimizar.

ORPAILLEUR

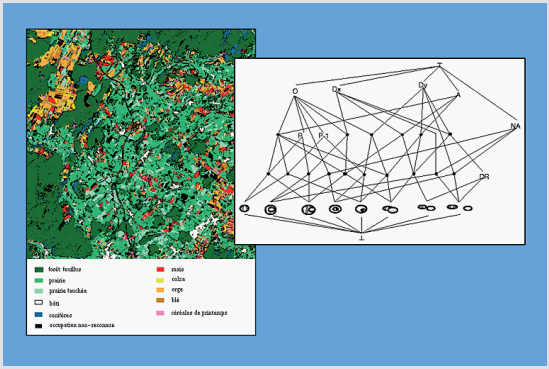
Descubrimiento del conocimiento, ingeniería del conocimiento

LORPAILLEUR es un equipo-proyecto del INRIA Nancy-Grand Est y LORIA desde principios de 2008. Es un equipo bastante grande y especial, ya que incluye informáticos, pero también un biólogo, químicos y un médico. Las ciencias de la vida, la química y la medicina son dominios de aplicación de primera importancia y el equipo desarrolla sistemas de trabajo para estos dominios.

El descubrimiento de conocimiento en bases de datos -en adelante KDD, Knowledge Discovery im Databases por su denominación inglesa- consiste

en procesar un gran volumen de datos para descubrir unidades de conocimiento que sean significativas y reutilizables. Si se equiparan las unidades de conocimiento con pepitas de oro y las bases de datos con terrenos o ríos por explorar, el proceso KDD puede compararse con la búsqueda de oro. Esto explica el nombre del equipo de investigación: en francés "orpailleur" se refiere a una persona que busca oro en ríos o montañas. Además, el proceso de KDD es iterativo, interactivo y generalmente controlado por un experto del dominio de los datos, llamado analista. El analista selecciona e interpreta un subconjunto de las unidades extraídas para obtener unidades de conocimiento que tengan una determinada plausibilidad. Como una persona que busca oro y que tiene un cierto conocimiento de la tarea y del lugar, el analista puede utilizar su propio conocimiento pero también el conocimiento sobre el dominio de los datos para mejorar el proceso KDD.

Una forma en la que el proceso KDD explota el conocimiento del dominio es vinculándolo a ontologías relacionadas con el dominio de los datos, como un paso hacia la noción de descubrimiento de conocimiento guiado por el conocimiento del dominio o KDDK. En el proceso KDDK, las unidades de conocimiento extraídas tienen todavía "una vida" después de la etapa de interpretación: se representan utilizando un formalismo de representación del conocimiento para ser integradas dentro de una ontología y reutilizadas para las necesidades de resolución de problemas. De este modo, el descubrimiento de conocimiento se utiliza para ampliar y actualizar las ontologías existentes, mostrando que el descubrimiento de conocimiento y la representación de conocimiento son tareas complementarias y dando cuerpo a la noción de KDDK .



Modelización de estructuras espaciales agrícolas extraídas de imágenes de satélite

Datos distribuidos por la red

Los datos distribuidos plantean problemas de rendimiento, como se muestra en la presentación de KERDATA más abajo. Pero hay una cuestión más fundamental relacionada con la privacidad. El aprendizaje colaborativo se ha desarrollado para satisfacer los requisitos de privacidad cuando se aprende con datos sensibles: la necesidad de garantizar "por diseño" un tratamiento compatible con el GDPR -sigla del inglés, General Data Protection Regulation- (por ejemplo, respetar la confidencialidad a las personas cuya imagen es captada por las cámaras).

KERDATA

Almacenamiento escalable para las Nubes y más allá.
La convergencia HPC-Big Data-AI y el continuo digital.

Las herramientas y las culturas de cómputo de alto rendimiento (HPC, o High Performance Computing) y de la analítica del Big Data han evolucionado de forma divergente. Esto va en detrimento de ambas. Sin embargo, los grandes cálculos generan Big Data y se necesitan potentes recursos informáticos para analizarlos. Más recientemente, el aprendizaje automático ha surgido con fuerza como un poderoso medio para permitir el análisis de datos relevantes a escala. Dado que la investigación científica depende cada vez más de la computación de alta velocidad y de la analítica de datos, la interoperabilidad potencial y la convergencia a escala de los ecosistemas correspondientes (HPC, Big Data, AI) es crucial para el futuro. En particular, un hito clave será lograr la convergencia mediante abstracciones y técnicas comunes para el almacenamiento y el procesamiento de datos en apoyo de flujos de trabajo complejos que combinen simulaciones, análisis y aprendizaje. Estos flujos de trabajo de aplicaciones necesitarán dicha convergencia para ejecutarse en infraestructuras híbridas que combinen sistemas HPC, nubes y dispositivos de proximidad, en un continuo digital completo.

Apoyar la IA en todo el espectro digital

Integrar y procesar de manera oportuna flujos de datos de alta frecuencia procedentes de múltiples sensores dispersos por un amplio territorio requiere técnicas y equipos de cómputo de alto rendimiento. Por ejemplo,

una solución de detección de terremotos basada en el aprendizaje automático tiene que diseñarse conjuntamente con expertos en informática distribuida y ciber-infraestructura para permitir las alertas en tiempo real. Debido al gran número de sensores y a su elevada frecuencia de muestreo, un enfoque centralizado tradicional que transfiera todos los datos a un único punto (por ejemplo, un sistema HPC o un centro de datos en la nube tradicional) puede resultar poco práctico. El equipo-proyecto KerData investiga soluciones innovadoras para el diseño de una arquitectura eficiente de procesamiento de datos en infraestructuras híbridas que combinan supercomputadoras, nubes y sistemas de proximidad, en apoyo del aprendizaje automático distribuido (y, en general, del análisis escalable de datos distribuidos).

En particular, basándose en los resultados anteriores del equipo en el área de los sistemas de procesamiento de flujos eficientes, el objetivo ahora es explorar enfoques para el almacenamiento de datos unificados, el procesamiento y el análisis basado en el aprendizaje automático en todo el continuo digital (es decir, para aplicaciones altamente distribuidas implementadas en infraestructuras híbridas de cercanía/nube/HPC). Las aplicaciones típicas incluyen flujos de trabajo complejos que combinan simulaciones y análisis, por ejemplo, gemelos digitales con mejora de datos.

Aprendizaje automático en el contexto del procesamiento de flujos de proximidad.

Esta reciente línea de investigación de Kerdata se desarrolla en estrecha colaboración con el grupo de la Manish Rutgers University, y con el equipo de LACODAM. Su objetivo es mejorar la exactitud de los sistemas de alerta temprana de terremotos (EEW) mediante el aprendizaje automático. Los sistemas de EEW están diseñados para detectar y caracterizar terremotos de mediana y gran magnitud antes de que sus efectos dañinos lleguen a un lugar determinado.

Los métodos tradicionales de EEW basados en sismómetros no logran identificar con exactitud los grandes terremotos debido a su sensibilidad a la velocidad del movimiento del suelo. Por otra parte, las estaciones de GPS de alta precisión, recientemente introducidas, no son eficaces para identificar terremotos de tamaño medio debido a su propensión a producir datos ruidosos. Además, las estaciones GPS y los sismógrafos pueden

desplegarse en grandes cantidades en diferentes lugares y producir un volumen significativo de datos, lo que afecta al tiempo de respuesta y a la solidez de los sistemas de alerta temprana.

En la práctica, los sistemas de alerta temprana pueden considerarse un típico problema de clasificación en el campo del aprendizaje automático: los datos de los multisensores se dan en la entrada, y la gravedad del terremoto es el resultado de la clasificación. En este artículo presentamos el sistema de Alerta Temprana de Terremotos con Sensores Múltiples Distribuidos (DMSEEW), un novedoso enfoque basado en el aprendizaje automático que combina datos de ambos tipos de sensores (estaciones GPS y sismómetros) para detectar terremotos de mediana y gran magnitud.

DMSEEW se basa en un nuevo método de apilamiento de ensambles o conjuntos que han sido evaluado en un conjunto de datos del mundo real validado con geocientíficos. El sistema se basa en una infraestructura distribuida geográficamente (que se puede implementar en nubes y sistemas de borde), lo que garantiza un cálculo eficiente en términos de tiempo de respuesta y robustez ante fallos parciales de la infraestructura. Nuestros experimentos muestran que DMSEEW es más exacto que el enfoque tradicional de sólo sismómetros y que el enfoque de sensores combinados (GPS y sismómetros) que adopta la regla de la fuerza relativa. Estos resultados han sido reconocidos por la comunidad internacional de IA mediante un "Outstanding Paper Award - Special Track on AI for Social Impact" en la AAAI-20, una conferencia "A*" en el área de la Inteligencia Artificial:

- Kévin Fauvel, Daniel Balouek-Thomert, Diego Melgar, Pedro Silva, Anthony Simonet, et al. A Distributed MultiSensor Machine Learning Approach to Earthquake Early Warning. AAAI 2020 - 34th AAAI Conference on Artificial Intelligence, Feb 2020, New York, United States. pp.1-9

Otros equipos de proyecto en este ámbito: MODAL (Lille), XPOP (Saclay).

5.2.3 Aprendizaje automático para la biología y la salud

En esta sección se enumeran cuatro equipos de proyectos que utilizan y desarrollan algunos aspectos del aprendizaje automático para problemas de Biología y Salud. Otros equipos pueden encontrarse en la sección de neurociencias y cognición.

Muchas de las aplicaciones del aprendizaje profundo se han destacado en la literatura (por ejemplo, en el libro de Eric Topo "Deep Medicine").

Las ciencias de la vida son uno de los campos más complicados, pero un campo de aplicación ideal: hay fuertes (y positivas) apuestas sociales y económicas, ya hay grandes cantidades de datos y conocimientos disponibles y formalizados. Al hablar de aplicaciones críticas para la vida, las exigencias son mucho más fuertes que en otros ámbitos en términos de verificación y validación, transparencia y trazabilidad, y explicabilidad, con el fin de generar confianza.

ABS

Algoritmos, biología, estructura

La biología computacional estructural (Computational structural biology por su denominación en inglés, o CSB) se ocupa de dilucidar la relación entre la estructura, la dinámica y las funciones de las biomoléculas. La CSB se nutre de datos experimentales de diversa índole. Por un lado, los proyectos de secuenciación del genoma dan acceso a las secuencias de proteínas, y se han archivado ~ 120 millones de secuencias en UNiProtKB/TrEMBL. Por otro lado, los experimentos de determinación de estructuras (especialmente la cristalografía de rayos X y la criomicroscopía electrónica) proporcionan modelos geométricos de moléculas; coordenadas atómicas. Por desgracia, solo se han resuelto ~ 150.000 estructuras. Con una estructura para ~ 1000 secuencias, apenas sabemos nada sobre las funciones biológicas a nivel atómico/molecular. Este escenario se debe a la alta dimensionalidad de los sistemas moleculares. Más concretamente, recordemos los tres siguientes factores.

En primer lugar, la conformación de una molécula con n átomos se caracteriza por 3n coordenadas cartesianas y 3n - 6 grados de libertad – es

necesario sacar el cálculo mediante movimientos rígidos. En la práctica, n ∈ [103,105].

En segundo lugar, a cada conformación se le asocia un entorno de energía potencial (PEL. o potential energy landscape). El PEL está definido por una función de R3n 7→ R, que es extremadamente compleja - el número de puntos críticos es exponencial en la dimensión.

En tercer lugar, las moléculas se deforman continuamente, y sus propiedades macroscópicas dependen de los valores medios del conjunto calculados sobre las regiones del PEL, como nos dice la física estadística. Por lo tanto, la estimación de las propiedades estructurales, termodinámicas y dinámicas son problemas muy difíciles

Resumiendo, hay tres retos principales en la CSB:

- Predecir la estructura tridimensional de una proteína a partir de su secuencia de aminoácidos. Este reto se investiga en el contexto del experimento bienal de toda la comunidad llamado "Evaluación Crítica de Predicción Estructural de las Proteínas" (CASP) -véase más abajo-.

- Estimar las propiedades termodinámicas y cinéticas de una proteína o un complejo proteico a partir de su estructura.

- Reconstruir la estructura de máquinas moleculares que integran hasta cientos de subunidades, un requisito previo para estudiar su función.

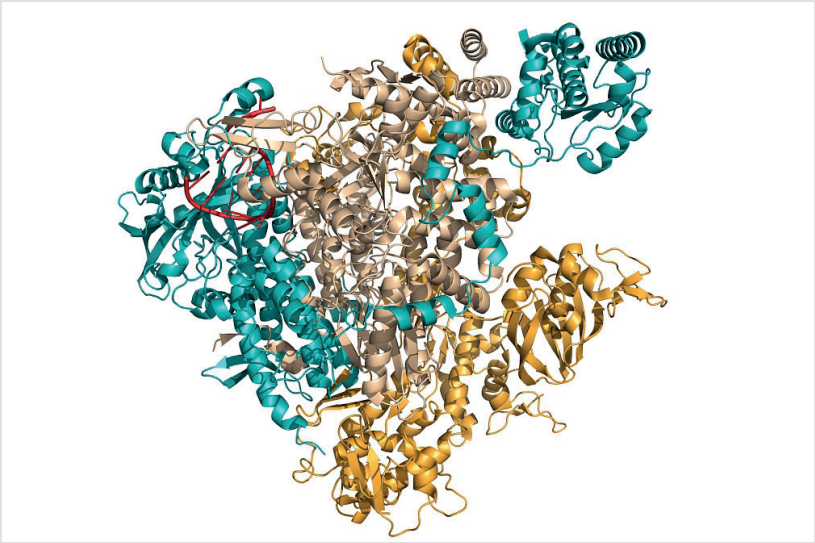
- El equipo-proyecto ABS desarrolla métodos originales para dilucidar estos problemas. Estos métodos recurren y contribuyen a varias disciplinas de las ciencias de la informática y las matemáticas aplicadas:

- La geometría y la topología, ya que los modelos estructurales son grafos incrustados en 3D.

- Optimización combinatoria, ya que los grafos son representaciones ubicuas tanto para las moléculas como para las redes moleculares.

- Aprendizaje automático, tanto supervisado (regresión, clasificación)

como no supervisado (agrupación, reducción de la dimensionalidad, matemáticas numéricas).



Modelado de la polimerasa del virus de la gripe - © INRIA

MIMESIS

Anatomía Computacional y Simulación Médica

MIMESIS desarrolla nuevas soluciones en el campo de la formación quirúrgica y las intervenciones asistidas por computadora para reducir el riesgo y mejorar las terapias guiadas por imágenes y señales.

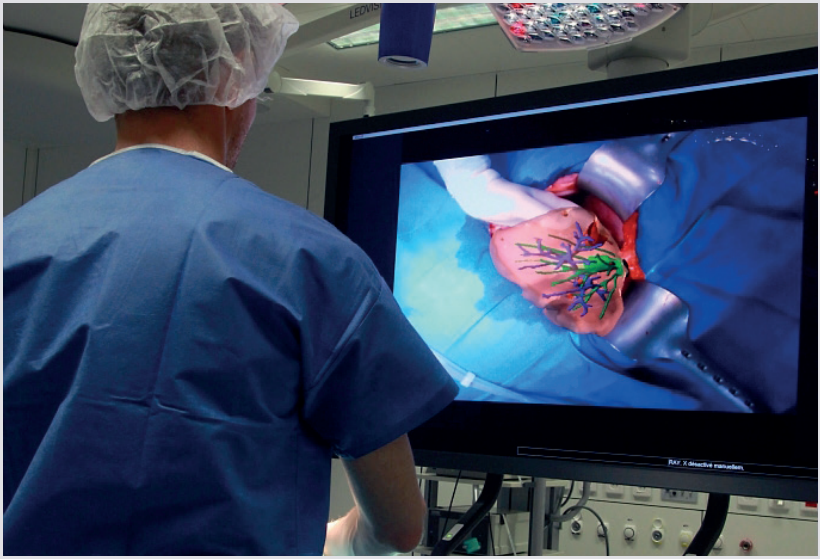
Modelos computacionales específicos para el paciente en tiempo real

Estamos desarrollando simulaciones computacionalmente eficientes, estables y precisas de (i) la deformación de los tejidos blandos y otros fenómenos biofísicos para proporcionar un feedback instantáneo y realidad aumentada durante la cirugía; (ii) la actividad eléctrica del cerebro y el comportamiento de los mamíferos para mejorar las terapias de

neuromodulación médica en los pacientes. Nuestra investigación también aborda la parametrización de modelos para describir las características específicas de los pacientes de (i) los tejidos blandos (forma, material, conductividad, etc.); (ii) las mediciones electromagnéticas de la actividad cerebral (electro/magnetoencefalografía, potenciales de campo local, actividad de una sola neurona). Por extensión, también desarrollamos modelos numéricos de las interacciones entre los tejidos y el instrumental, un componente clave de los sistemas de entrenamiento quirúrgico.

Simulación basada en datos

Esta línea de investigación tiene como objetivo tender puentes entre la imagen médica y la práctica clínica mediante la adopción de datos preoperatorios en el momento de la intervención. Para ello combinamos métodos bayesianos con técnicas avanzadas basadas en la física para manejar las incertidumbres en las simulaciones basadas en señales e imágenes. También estamos desarrollando redes neuronales que pueden predecir la compleja física de los tejidos blandos y combinarlos con los métodos clásicos para garantizar la explicabilidad y la exactitud de la predicción.



Intervención asistida por computadora - © Inria _ Photo F. Nussbaumer - Signatures

MONC

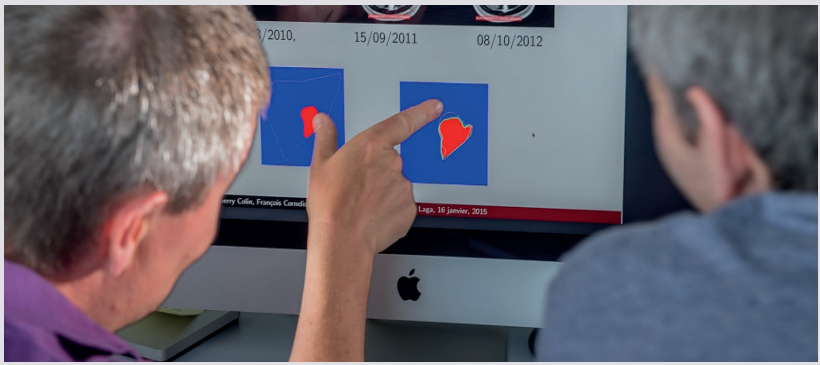
Modelización matemática para la oncología

El equipo-proyecto Monc trabaja en el campo de la medicina basada en datos contra el cáncer. Combinamos los modelos matemáticos y la IA con datos para abordar retos relevantes para biólogos y clínicos.

Sus objetivos son los siguientes:

- Mejorar nuestra comprensión de la biología y la farmacología del cáncer,
- Ayudar al desarrollo de nuevos enfoques terapéuticos,
- Desarrollar herramientas personalizadas para la toma de decisiones para el seguimiento de la enfermedad y la evaluación de terapias.

Más concretamente, estamos desarrollando modelos matemáticos -que implican ecuaciones en derivadas parciales (EDP) y que se construyen a partir de un conocimiento biológico y médico preciso - combinados con novedosas técnicas de asimilación de datos, procesamiento de imágenes, métodos estadísticos e inteligencia artificial (aprendizaje automático, aprendizaje profundo) - para construir herramientas numéricas basadas en los datos cuantitativos disponibles sobre el seguimiento del cáncer.



Modelización matemática para la oncología – Predicción del crecimiento del tumor y estimación de la respuesta al tratamiento - © Inria _ Photo H. Raguet

Cada tipo de cáncer es diferente y los modelos se enfocan específicamente a un número limitado de patologías (*por ejemplo, metástasis cerebrales y pulmonares, meningioma, gliomas, sarcoma de tejidos blandos, tumores de pulmón).

SISTM

Estadísticas en Biología de Sistemas y Medicina Traslacional

SISTM son las siglas de Statistics in Systems Biology and Translational Medicine. La investigación realizada en este equipo se aplica al campo de las ciencias médicas y, más concretamente, a las enfermedades infecciosas y la inmunología. Se requieren métodos específicos para tratar los datos de alta dimensión que se generan en este campo. En concreto, las mejoras biotecnológicas permiten medir los distintos tipos de células y su actividad de forma mucho más precisa. Así, en una sola muestra de sangre de un determinado paciente, se pueden determinar potencialmente millones de tipos de células (2^40) por citometría de masas, la expresión de 20.000 genes por secuenciación de ARN y la producción de cientos o miles de proteínas por multiplexación o espectrometría. Por lo tanto, el análisis de estos datos requiere enfoques de reducción de dimensión (1,2), no supervisados (3), o supervisados (por ejemplo, basados en la técnica "Random Forest") (4), clasificación en un espacio multidimensional, pruebas estadísticas adaptadas para un entorno de alta dimensión (5). Los resultados obtenidos en estos espacios de alta dimensión proporcionan mucho más conocimiento a partir de estudios clínicos individuales, lo que resulta muy útil para el desarrollo de vacunas, por ejemplo (6). El siguiente paso es la adaptación de las intervenciones en función de los datos recogidos a lo largo de los ensayos (7).

1. Sutton M, Thiébaut R, Liquet B. Sparse partial least squares with group and subgroup structure. Stat Med (2018) 37:3338–3356. doi:10.1002/sim.7821

2. Lorenzo H, Misbah R, Odeber J, Morange PE, Saracco J, Tregouet DA, Thiebaut R. High-dimensional multi-block analysis of factors associated with thrombin generation potential. En Proceedings - IEEE Symposium on Computer-Based Medical Systems (Institute of Electrical and Electronics Engineers Inc.), 453–458. doi:10.1109/CBMS.2019.00094

3. Hejblum BP, Alkhassim C, Gottardo R, Caron F, Thiébaut R. Sequential dirichlet process mixtures of multivariate skew t-distributions for model-based clustering of flow cytometry data. Ann Appl Stat (2019) 13:638–660. doi:10.1214/18AOAS1209

4. Capitaine L, Genuer R, Thiébaut R. Fr\'echet random forests. (2019) Disponible en: http://arxiv.org/abs/1906.01741 [Consultado el 4 de junio de 2020]
5. Agniel, Denis, Hejblum B. Variance component score test for time-course gene set analysis of longitudinal RNA-seq data | Biostatistics | Oxford Academic. Disponible en: https://academic.oup.com/biostatistics/article/18/4/589/3065599 [Consultado el 5 de junio de 2020]
6. Rechtien A, Richert L, Lorenzo H, Martrus G, Hejblum B, Dahlke C, Kasonta R, Zinser M, Stubbe H, Matschl U, et al. Systems Vaccinology Identifies an Early Innate Immune Signature as a Correlate of Antibody Responses to the Ebola Vaccine rVSV-ZEBOV. Cell Rep (2017) 20:2251–2261. doi:10.1016/j.celrep.2017.08.023
7. Pasin C, Dufour F, Villain L, Zhang H, Thiébaut R. Controlling IL-7 Injections in HIV-Infected Patients. Bull Math Biol (2018) 80:2349–2377. doi:10.1007/s11538-018-0465-8

5.2.4 Acciones exploratorias (AEx) y retos de Inria

Desafío Inria – "Enfoques híbridos para la inteligencia artificial explicable" (HyAIAI)

Equipos de proyecto: LACODAM, TAU, SCOOL, MAGNET, ORPAILLEUR, MULTISPEECH

Existe una tendencia de investigación emergente cuyo objetivo es proporcionar interpretaciones para la decisión de los algoritmos de "caja negra" de ML como los de Aprendizaje Profundo (DL).

En el Desafío Inria de HyAIAI, afirmamos que es necesaria una comunicación bidireccional entre un modelo de DL y un usuario: por supuesto, el usuario debe entender las decisiones de DL, pero cuando el usuario participa en el entrenamiento del modelo de DL, también debe ser capaz de proporcionar una retroalimentación expresiva al modelo. Creemos que esta comunicación bidireccional requiere un enfoque híbrido: los modelos numéricos complejos deben desempeñar el papel de motor de aprendizaje debido a su rendimiento, pero deben combinarse con modelos simbólicos para garantizar una comunicación eficaz con el usuario.

Desafío Inria – "Cómputo de alto rendimiento y Big Data (HPC-BigData)"

Consulte la lista completa de equipos de proyecto en https://project.inria.fr/hpcbigdata/.

El análisis de Big Data es cada vez más intensivo en términos del cómputo gracias al aprendizaje profundo, mientras que el manejo de datos se está convirtiendo en una preocupación importante para la computación científica. El reto HPC-Big Data reúne a equipos de las áreas de HPC, Big Data y Machine Learning para trabajar en la intersección entre estos dominios.

AEx-AI4HI - Inteligencia artificial para la inteligencia humana

Equipo-proyecto: CORSE

El objetivo de AI4HI es aunar los avances en Inteligencia Artificial (clasificación, enfoques estadísticos, aprendizaje profundo) y las habilidades de compilación y enseñanza para poder mejorar la enseñanza mediante la generación automática de ejercicios y su recomendación a los estudiantes. El proyecto se centra en la enseñanza en programación y depuración a nivel de principiantes.

AEx-MALESI – Aprendizaje automático para la simulación

Equipo-proyecto: TONUS

Las simulaciones físicas requieren una resolución ultraprecisa de las ecuaciones diferenciales parciales (EDP). Los esquemas numéricos actuales pueden generar una importante contaminación numérica. El proyecto pretende desarrollar métodos de aprendizaje basados en imágenes para corregir estas deficiencias numéricas, al tiempo que se demuestran las importantes propiedades de convergencia y universalidad.

AEx-SR4SG : Aprendizaje colaborativo secuencial de recomendaciones para la jardinería sostenible

Equipo-proyecto: SCOOL

El objetivo del SR4SG es doble: por un lado agrupar una ambiciosa comunidad mixta en torno al tema “Aprendizaje de refuerzo para la jardinería sostenible “ y por otro lado proporcionar una plataforma de aplicación común para integrar progresivamente los conocimientos de investigación de todas las partes interesadas (aprendizaje secuencial, ontología, hci, cómputo distribuido, certificación de datos, botánica, ecología funcional, epidemiología, agronomía, agroecología, etc.).

AEx-TRACME - Rutas causales multiescala

Equipo-proyecto: GEOTSTAT

Este proyecto se centra en la modelización de un sistema físico a partir de mediciones sobre dicho sistema. ¿Cómo, a partir de las observaciones, construir un modelo fiable de la dinámica del sistema? Cuando varios procesos interactúan a diferentes escalas, ¿cómo obtener un modelo significativo para cada una de ellas? ¿Cómo relacionar estos modelos con cantidades físicas, como la cantidad de energía, o la de información, que se procesan a cada escala? Este proyecto propone identificar clases de estados de sistema causalmente equivalentes, y luego modelar su evolución con un proceso estocástico. Es necesario renormalizar estas ecuaciones para poder establecer una conexión entre la escala del continuo y aquella, arbitraria, a la que se adquieren los datos. Las aplicaciones involucran principalmente las ciencias naturales.

AEx-FLAMED - Aprendizaje colaborativo y análisis de datos médicos

Equipo-proyecto: MAGNET

FLAMED tiene como objetivo explorar un enfoque descentralizado de la Inteligencia Artificial aplicada a la salud. En estrecha colaboración con el hospital universitario de Lille, el objetivo de FLAMED es llevar a cabo tareas de análisis de datos y aprendizaje automático (aprendizaje agregado descentralizado) en las que participen varios hospitales, permitiendo que cada centro conserve sus datos internamente y garantizando la confidencialidad.

AEx-MAMMALS - Modelos aumentados por la memoria para un servicio de aprendizaje automático de baja latencia

Equipo-proyecto: NEO

MAMMALS tiene como objetivo proporcionar inferencias de baja latencia mediante la ejecución - cerca del usuario final - de modelos simples de aprendizaje automático que también pueden beneficiarse de un (pequeño) almacén de datos local de ejemplos. El énfasis está en los algoritmos para aprender en línea qué almacenar localmente para mejorar la calidad de la inferencia y lograr la adaptación de dominio. MAMALS nos ayudará a profundizar en la comprensión de la

relación entre la memorización y la generalización, que sigue siendo insuficiente incluso en el entorno estático.

5.2.5 Software: SCIKIT-LEARN

La biblioteca de referencia de Python para aprendizaje automático

A nivel mundial, scikit-learn es el primer software de aprendizaje automático de código abierto dirigido por una comunidad de investigadores. Rivaliza en popularidad con las herramientas desarrolladas por el GAFA (acrónimo para el conjunto de google, Apple, Facebook y Amazon).

La visión de scikit-learn.

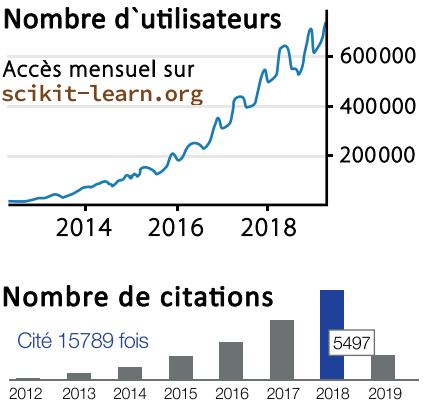
El equipo de Inria Parietal lleva desarrollando scikit-learn desde 2010 con el fin de facilitar el acceso al aprendizaje estadístico al mayor número de personas posible, especialmente a los neurocientíficos. Al proporcionar una herramienta eficaz, sencilla de utilizar y muy bien documentada con cientos de ejemplos, los desarrolladores de scikit-learn han contribuido a la democratización del aprendizaje estadístico que ha impulsado la actual revolución de la inteligencia artificial. Con un impacto mucho más amplio que el de las neurociencias, los investigadores e ingenieros de Inria que están detrás del éxito de scikit-learn han permitido el uso del aprendizaje estadístico en todas las ciencias experimentales como la química, la biología y la física, así como en muchas aplicaciones industriales.

Scikit-learn: una referencia en el aprendizaje estadístico.

Scikit-learn reúne más de 180 modelos diferentes de aprendizaje estadístico. Abarca muchos aspectos de esta disciplina de la matemática aplicada y proporciona un conjunto de herramientas algorítmicas de referencia, como las que se encuentran en los libros sobre el tema. Su documentación –http://scikit-learn.org– es en sí misma una introducción al aprendizaje estadístico. Se considera una herramienta pedagógica y tendría más de mil páginas en formato papel. Scikit-learn no incluye directamente arquitecturas de aprendizaje profundo, pero puede conectarse a bibliotecas de DL según sea necesario.

Métricas de uso.

Debido a que scikit-learn es un software libre, es difícil tener cifras exactas del número de usuarios. Ahora bien, las estadísticas del sitio web indicaron más de 42 millones de visitas en 2018 y 700.000 usuarios mensuales (figura de la derecha).



GitHub, que alberga el código fuente del proyecto, informa de cerca de 17.000 ramificaciones y 35.000 estrellas. Scikit-learn representa 39 años*persona de trabajo. Es el tercer software de aprendizaje automático de código abierto más popular, por detrás de dos herramientas de software desarrolladas por Google. Una encuesta realizada hace unos años identificó que el 63% de los usuarios pertenecían a la industria, y el 34% al mundo académico. El artículo académico de referencia ha sido citado 25.000 veces en Google Scholar desde 2012, con 8200 citas en 2019 (figura de la derecha).

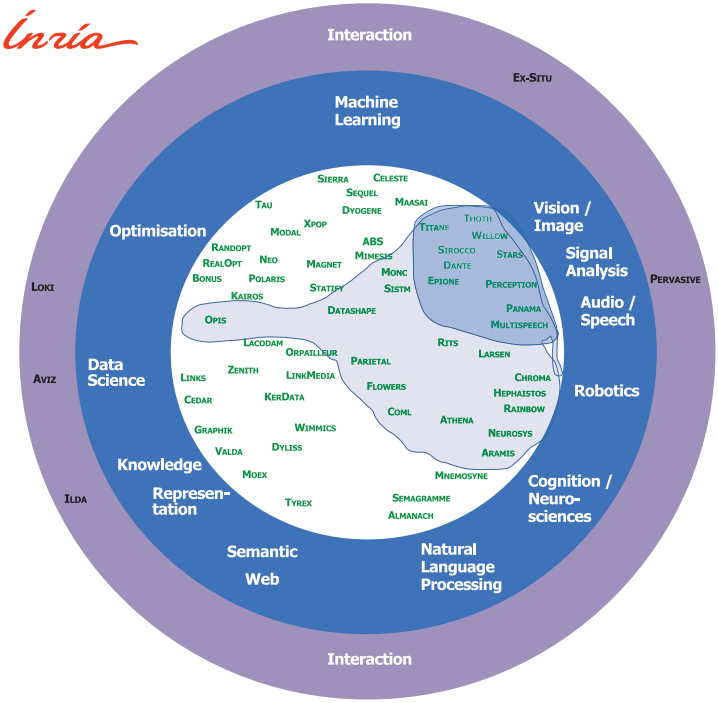
El consorcio scikit-learn alojado en la fundación Inria nació en septiembre de 2018 con el apoyo de 7 empresas: Microsoft, BCG, AXA, BNP Paribas-Cardif, Intel, NVIDIA y Dataiku, a las que se sumó Fujitsu. Esta asociación/patrocinio demuestra el impacto industrial de scikit-learn y permitirá la financiación a largo plazo del software.

5.3 Análisis de señales, visión, habla

El análisis de señales, en particular la visión y el reconocimiento de patrones, es el punto de partida del actual revuelo del aprendizaje profundo: desde 2012, los sistemas de aprendizaje profundo ganaron "todos" los retos en visión y reconocimiento de patrones, algo que convenció a casi todos los investigadores y profesionales del campo para pasarse al aprendizaje profundo. Estos éxitos también llegaron al reconocimiento del habla, y poco a poco se hicieron muy populares en la mayoría de los campos de las ciencias de la informática, transfiriéndose rápidamente a la industria correspondiente: el sistema MobilEyevision

potencia las capacidades de autoconducción de los coches, mientras que los asistentes de voz como Siri, Cortana o Amazon Echo son utilizados cada día por millones de usuarios.

El reconocimiento de objetos -o, en un sentido más amplio, la comprensión de escenas- es el último reto científico de la visión por computadora: Tras 40 años de investigación, aunque se han hecho enormes progresos en la identificación de objetos familiares (silla, persona, mascota), categorías de escenas (playa, bosque, oficina) y patrones de actividad (conversación, baile, picnic) representados en fotos familiares, segmentos de noticias o largometrajes, la comprensión de escenas completas a nivel de los humanos sigue estando muy lejos de las capacidades de los sistemas de visión actuales, en parte debido a la falta de sentido común (es decir, conocimiento general a priori) de todos los sistemas de aprendizaje actuales. Sin embargo, el impacto de la tecnología actual y futura de reconocimiento de objetos y comprensión de escenas seguirá creciendo en ámbitos de aplicación tan variados como la defensa, el entretenimiento, la atención sanitaria, la interacción persona-computadora, la recuperación de imágenes y la minería de datos, la robótica industrial y personal, la manufactura, el análisis científico de imágenes, la vigilancia y la seguridad, y el transporte.



Los retos del **análisis de señales para la visión** son (i) la escalabilidad; (ii) pasar de imágenes fijas a vídeo; (iii) la multimodalidad; (iv) la introducción de conocimientos a priori.

Escalabilidad

Los sistemas de visión modernos deben ser capaces de tratar un gran volumen de datos y una alta frecuencia en el momento de la inferencia: por ejemplo, los sistemas de vigilancia en lugares públicos, los robots que se mueven en entornos desconocidos o los motores de búsqueda de imágenes en la web tienen que procesar enormes cantidades de datos. Los sistemas de visión no sólo deben procesar estos datos a gran velocidad, sino que deben alcanzar altos niveles de precisión para así poder liberar a los operadores de la comprobación de los resultados y el post-procesamiento. Incluso unos índices de precisión del 99,9% para la clasificación de imágenes en operaciones de misión crítica no son suficientes cuando se procesan millones de imágenes, ya que el 0,1% restante necesitará horas de procesamiento humano.

De la imagen al vídeo

A pesar de las limitaciones de la tecnología actual de comprensión de escena, en los últimos diez años se han realizado enormes progresos, debido en parte a la formulación del reconocimiento de objetos como un problema de búsqueda de patrones estadísticos. En general, se hace hincapié en las características que definen los patrones y en los algoritmos utilizados para aprenderlos y reconocerlos, más que en la representación de las categorías de objetos, escenas y actividades, o en la interpretación integrada de los distintos elementos de la escena.

Multimodalidad

La comprensión de los datos visuales puede mejorarse por diferentes medios: en la red, los metadatos proporcionados con las imágenes y los vídeos pueden utilizarse para eliminar ciertas suposiciones y guiar al sistema hacia el reconocimiento de objetos, eventos o situaciones específicas. Otra opción es utilizar la multimodalidad, es decir, señales procedentes de varios canales, por ejemplo, infrarrojos, láser, datos magnéticos, etc. También es deseable utilizar una combinación de señal auditiva con visión (imágenes o vídeo) en caso de estar disponible.

Introducción de conocimiento a priori

Otra opción para mejorar las aplicaciones de visión es introducir conocimientos a priori en el motor de reconocimiento. Un ejemplo consiste en añadir información sobre la anatomía y la patología de un paciente para un mejor análisis de las imágenes biomédicas; en otros ámbitos, la información contextual, la información sobre una situación, sobre una tarea, los datos de localización, etc. pueden utilizarse para eliminar la ambigüedad sobre las posibles interacciones. Sin embargo, la cuestión de cómo proporcionar este conocimiento a priori no está resuelta en términos generales: hay que establecer métodos y representaciones de conocimiento específicos para tratar una aplicación determinada en la comprensión de la visión.

WILLOW

Modelos de reconocimiento visual de objetos y comprensión de escenas

WILLOW aborda problemas fundamentales de la visión por computadora, como la percepción tridimensional, la fotografía computacional y la comprensión de imágenes y vídeos. Investiga nuevos modelos del contenido de las imágenes (¿en qué consiste un buen vocabulario visual?) y de procesos de interpretación (¿qué es una buena arquitectura de reconocimiento?).

A pesar de los enormes avances logrados en el reconocimiento visual en los últimos 10 años, los sistemas actuales de reconocimiento visual siguen necesitando grandes cantidades de datos de entrenamiento cuidadosamente anotados, suelen utilizar arquitecturas de caja negra que no modelan la naturaleza física tridimensional del mundo visual y no captan la semántica del mundo real. WILLOW aborda estas limitaciones mediante el desarrollo de modelos de todo el proceso de comprensión visual que se pueden conseguir sin necesidad de supervisión directa, permiten un razonamiento complejo sobre los datos visuales y se basan en las interacciones con el mundo físico. Más concretamente, WILLOW aborda retos científicos fundamentales en cuatro ejes de investigación: (i) el reconocimiento visual en imágenes y vídeos con énfasis en el aprendizaje débilmente supervisado; (ii) el aprendizaje de representaciones visuales incorporadas para la manipulación y locomoción robótica; (iii) la restauración y mejora de imágenes; y (iv) el

modelado, análisis y recuperación de objetos y escenas en 3D.

Sus últimos logros incluyen trabajos teóricos sobre los fundamentos geométricos de la visión por computadora, nuevos avances en tareas de restauración de imágenes como eliminación de borrosidad, reducción del ruido o el remuestreo (upsampling), y métodos débilmente supervisados de aprendizaje de representaciones efectivas para la recuperación de textos y vídeos y la localización temporal de acciones. Los miembros de WILLOW trabajan en estrecha colaboración con los equipos SIERRA y THOTH de Inria, así como con investigadores de la Universidad Carnegie-Mellon, la UC Berkeley y el laboratorio de IA de Facebook, colaboraciones que ponen de manifiesto la fuerte sinergia entre el aprendizaje automático y la visión por computadora, con nuevas oportunidades en campos que van desde la arqueología a la robótica. Los retos futuros incluyen el desarrollo de modelos mínimamente supervisados para el reconocimiento visual en conjuntos de datos de imágenes y vídeos a gran escala, y agentes autónomos impulsados por la visión.

SFNET: Aprendiendo flujo semántico basado en el reconocimiento de objetos

STARS

Sistemas de reconocimiento de actividad espacio-temporal

En los últimos años, se han realizado muchos estudios avanzados en el campo de Visión por Computadora y, en particular, en el de la Comprensión de Escenas. La comprensión de la escena es el proceso,

a menudo en tiempo real, de percibir, analizar y elaborar una interpretación de una escena 3D dinámica observada a través de una red de sensores (por ejemplo, cámaras de vídeo). Este proceso consiste esencialmente en asociar la información de las señales procedentes de los sensores que observan la escena con los modelos que los humanos utilizan para comprenderla. Por lo tanto, la comprensión de la escena consiste en añadir y extraer la semántica de los datos de los sensores que caracterizan una escena. Esta escena puede contener una serie de objetos físicos de diversos tipos (por ejemplo, personas, vehículos) que interactúan entre sí o con un entorno (por ejemplo, equipos) más o menos estructurado. La escena puede durar unos instantes (por ejemplo, la caída de una persona) o unos meses (por ejemplo, la depresión de una persona), y puede limitarse a un portaobjetos de laboratorio observado a través de un microscopio o superar el tamaño de una ciudad. Los sensores suelen ser cámaras (por ejemplo, omnidireccionales, de infrarrojos, de profundidad), pero también pueden ser micrófonos u otros tipos de sensores (por ejemplo, células ópticas, sensores de contacto, sensores fisiológicos, acelerómetros, radares, detectores de humo, teléfonos inteligentes).

La comprensión de la escena se inspira en la visión cognitiva y requiere la combinación de al menos tres campos: **la visión por computadora, el aprendizaje automático y la ingeniería de software**. La comprensión de la escena puede alcanzar cinco niveles de funcionalidad genérica de la visión por computadora: detección, localización, seguimiento, reconocimiento y comprensión. Pero los sistemas de comprensión de la escena van más allá de la detección de características visuales como esquinas, bordes y regiones en movimiento para extraer información relacionada con el mundo físico que sea significativa para los operadores humanos. La comprensión de las escenas también requiere capacidades de visión por computadora más sólidas, robustas y flexibles, dotándolas de competencia cognitiva: la capacidad de aprender, adaptarse, sopesar soluciones alternativas y desarrollar nuevas estrategias de análisis e interpretación.

En cuanto a la comprensión de la escena, el equipo de STARS ha desarrollado sistemas automatizados originales para comprender los comportamientos humanos en una gran variedad de entornos para diferentes aplicaciones:

- En las estaciones de metro, en las calles y a bordo de los trenes: peleas, equipajes abandonados, graffiti, fraudes, comportamiento de las multitudes,

- En las plataformas de los aeropuertos: aterrizajes de aeronaves, reabastecimiento de combustible de aeronaves, carga/descarga de equipajes, señalamiento (marshalling),

- En agencias bancarias: asalto a bancos, control de acceso en edificios, uso de cajeros automáticos,

- Aplicaciones de atención domiciliaria para vigilar las actividades de las personas mayores: cocinar, dormir, preparar el café, ver la televisión, preparar el pastillero, caerse;

- En la casa inteligente, la vigilancia del comportamiento en la oficina para la inteligencia ambiental: leer, beber;

- Supervisión del mercado para la inteligencia empresarial: detenerse ante un producto, hacer cola, retirar artículos;

- Aplicaciones biológicas: seguimiento de avispas;

- Biometría: expresión facial; y

- Demencia y trastorno cognitivo: diagnóstico precoz basado en el seguimiento del comportamiento y las emociones

Para desarrollar estos sistemas, el equipo de STARS diseñó nuevas tecnologías para la generación de vídeo [Wang 2020], la re-identificación de personas [Chen 2021] y para el reconocimiento de actividades humanas, en particular utilizando en cámaras de vídeo 2D o 3D. Más concretamente, los investigadores combinaron 4 categorías de algoritmos para reconocer actividades humanas:

- Motores de reconocimiento que utilizan ontologías expresadas por reglas que representan el conocimiento del experto. Estos motores de reconocimiento de actividades pueden ampliarse fácilmente y permiten la integración posterior de información adicional de los sensores cuando está disponible [Crispim 2016].

- Métodos de aprendizaje supervisado basados en muestras positivas/negativas representativas de las actividades objetivo especificadas por los usuarios. Estos métodos suelen basarse en el aprendizaje profundo que computa descriptores espacio-temporales robustos [Das 2019].

- Métodos de aprendizaje no supervisados (totalmente automatizados o débilmente o parcialmente supervisados) basados en la agrupación de patrones de actividad frecuentes en grandes conjuntos de datos que pueden generar/descubrir nuevos modelos de actividad [Negin 2019].

- Mecanismos de atención (autosupervisión o enfoque en la dimensión espacial o temporal) para guiar los métodos de aprendizaje a centrarse en la información más destacada dentro de un vídeo [Das 2020].

C. Crispim-Junior, K. Avgerinakis, V. Buso, G. Meditskos, A. Briassouli, J. Benois-Pineau, Y. Kompatsiaris and F. Bremond. Semantic Event Fusion of Different Visual Modality Concepts for Activity Recognition, Transactions on Pattern Analysis and Machine Intelligence - PAMI 2016.

S. Das, R. Dai, M. Koperski, L. Minciullo, L. Garattoni, F. Bremond and G. Francesca. Toyota Smarthome: Real-World Activities of Daily Living with supplementary. En Proceedings of the 17th International Conference on Computer Vision, ICCV 2019, en Seoul, Corea, del 27 de octubre al 2 de noviembre de 2019.

F. Negin and F. Bremond. An Unsupervised Framework for Online Spatiotemporal Detection of Activities of Daily Living by Hierarchical Activity Models, en Sensors 2019, 19, 1-27, doi:10.3390/s19194237; 29 de septiembre de 2019.

Y. Wang, P. Bilinski, F. Bremond and A. Dantcheva. G^{3}AN: Disentangling appearance and motion for video generation. En Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle-online, US, del 14 al 19 de junio de 2020.

S. Das, S. Sharma, R. Dai, F. Bremond and M. Thonnat. VPN: Learning Video-Pose Embedding for Activities of Daily Living. En Proceedings of the 16th European Conference on Computer Vision, ECCV 2020, arXiv:2007.03056, online, UK, del 23 al 28 de Agosto de 2020.

H. Chen, B. Lagadec and F. Bremond. Enhancing Diversity in Teacher-Student Networks via Asymmetric branches for Unsupervised Person Re-identification. En Proceedings of the IEEE Winter Conference on Applications of Computer Vision, WACV 2021, Virtual, del 5 al 9 de enero de 2021.

THOTH

Aprendizaje de modelos visuales a partir de datos a gran escala

La cantidad de imágenes y vídeos digitales disponibles en línea sigue creciendo a una velocidad fenomenal: los usuarios domésticos ponen sus películas en YouTube y sus imágenes en Flickr; los periodistas y científicos crean páginas web para difundir noticias y resultados de investigación y el sector público ahora cuenta con acceso a los archivos audiovisuales de las emisiones de televisión. En 2021, se preveía que casi el 82% del tráfico de Internet sería atribuible a vídeos y que un individuo tardaría más de 5 millones de años en ver la cantidad de vídeos que cruzarían las redes IP mundiales cada mes para entonces. Por tanto, la necesidad de anotar y realizar una indexación de estos contenidos visuales para usuarios domésticos y profesionales es apremiante y, de hecho, creciente. Los metadatos de texto y audio disponibles no suelen ser suficientes por sí solos para responder a la mayoría de las consultas y los datos visuales deben entrar en juego. Por otra parte, es inconcebible aprender los modelos de contenido visual necesarios para responder a estas consultas, anotando manualmente y con precisión cada concepto, objeto, escena o categoría de acción pertinente en una muestra representativa de condiciones cotidianas, aunque sólo sea porque puede ser difícil, o incluso imposible, decidir a priori cuáles son las categorías relevantes y el nivel de granularidad adecuado. El objetivo principal de THOTH es explorar automáticamente grandes colecciones de datos, seleccionar la información relevante y aprender la estructura y los parámetros de los modelos visuales. Hay tres retos principales: (1) el diseño y el aprendizaje de modelos estructurados capaces de representar información visual compleja; (2) el aprendizaje conjunto en línea de modelos visuales a partir de anotaciones textuales, sonido, imagen y vídeo; y (3) el aprendizaje y la optimización a gran escala. Otro aspecto importante es (4) la recogida y evaluación de datos.

La tecnología actual de reconocimiento de objetos y comprensión de escenas funciona en un entorno muy diferente; en su mayor parte se basa en motores de clasificación totalmente supervisados, y los modelos visuales son esencialmente plantillas rígidas (por partes) aprendidas a partir de imágenes etiquetadas a mano. La magnitud de los datos en línea y la naturaleza de las anotaciones incorporadas exigen apartarse

de este escenario totalmente supervisado. La idea principal del equipo-proyecto THOTH es desarrollar un nuevo marco para el aprendizaje de la estructura y los parámetros de los modelos visuales mediante la exploración activa de grandes fuentes de imágenes y vídeos digitales (archivos fuera de línea y contenidos en línea cada vez mayores, con millones de imágenes y miles de horas de vídeo), y la explotación de la débil señal de supervisión proporcionada por los metadatos que los acompañan. Este enorme volumen de datos visuales de entrenamiento nos permitirá aprender modelos no lineales complejos con un gran número de parámetros, como las redes convolucionales profundas y los modelos gráficos de orden superior. Se trata de un objetivo ambicioso, dado el enorme volumen y la variabilidad intrínseca de los datos visuales disponibles en línea y la falta de un formalismo universalmente aceptado para modelarlos. No obstante, la potencial recompensa es un gran avance en cuanto al reconocimiento de objetos visuales y la capacidad de comprensión de escenas. Además, los recientes avances a menor escala sugieren que esto es viable. Por ejemplo, ya es posible determinar la identidad de varias personas a partir de imágenes de noticias y sus pies de foto o aprender modelos de acción humana a partir de guiones de vídeo. Asimismo, ha habido avances recientes en la adaptación de

Aprendizaje de patrones de movimiento en vídeos

la tecnología de aprendizaje automático supervisado a entornos a gran escala, donde los datos de entrenamiento son muy grandes y potencialmente infinitos y algunos de ellos pueden no estar etiquetados. A su vez, han surgido métodos que adaptan la estructura de los modelos visuales a los datos, y la creciente potencia de cálculo y capacidad de almacenamiento de los computadores modernos son factores que, por supuesto, no deberían ignorarse.

SIROCCO

Representación, compresión y comunicación de datos visuales

El programa de investigación del equipo SIROCCO consiste en el diseño de modelos matemáticos y algoritmos para la obtención de imágenes computacionales, aprovechando los métodos de procesamiento de señales y aprendizaje automático, con un enfoque reciente en modalidades emergentes como las imágenes de alto rango dinámico, los campos de luz y las imágenes omnidireccionales. Los problemas de investigación que aborda el equipo se encuentran en la intersección entre el procesamiento de señales, la visión por ordenador, el aprendizaje automático y la teoría de la información. Los temas de investigación más precisos son:

- Análisis de datos visuales con problemas de visión por ordenador como la estimación de la profundidad y el flujo de la escena

- Métodos de procesamiento de señales y aprendizaje para la representación de datos visuales y compresión. Esto incluye modelos dispersos, de bajo rango y basados en grafos para diferentes modalidades de imagen

- Algoritmos para problemas inversos en el procesamiento de datos visuales como adquisición compresiva, restauración y superresolución.

- Herramientas de teoría de la información y codificación para la comunicación interactiva

Aprendizaje de la profundidad de la escena a partir de un subconjunto flexible de vistas de campo de luz densas y dispersas

EPIONE

E-Paciente: Imágenes, datos y modelos para la medicina Electrónica

El objetivo a largo plazo de EPIONE es contribuir al desarrollo de lo que se denomina el e-paciente (e-patient o paciente digital) para la e-medicine (medicina digital).

- El **e-paciente** (e-patient o paciente digital) es un conjunto de modelos computacionales del cuerpo humano capaces de describir y simular la anatomía y la fisiología de los órganos y tejidos del paciente, a varias escalas, para un individuo o una población. El e-paciente puede verse como un marco para integrar y analizar de forma coherente la información heterogénea medida sobre el paciente a partir de fuentes dispares: imágenes, fuentes biológicas, clínicas, sensores, etc.

- La **e-medicina** (o medicina digital) se define como las herramientas informáticas aplicadas al e-paciente para ayudar al médico y al cirujano en su práctica médica, para evaluar el diagnóstico/pronóstico y para planificar, controlar y evaluar la terapia.

Los modelos que rigen los algoritmos diseñados para los e-pacientes y la e-medicina proceden de diversas disciplinas: informática, matemáticas, medicina, estadística, física, biología, química, etc. Los parámetros de esos modelos deben ajustarse a un individuo o a una población en

función de las imágenes, señales y datos disponibles. Este ajuste se denomina personalización y suele requerir la resolución de difíciles problemas inversos.

Los objetivos de investigación de EPIONE se organizan en torno a 5 ejes científicos:

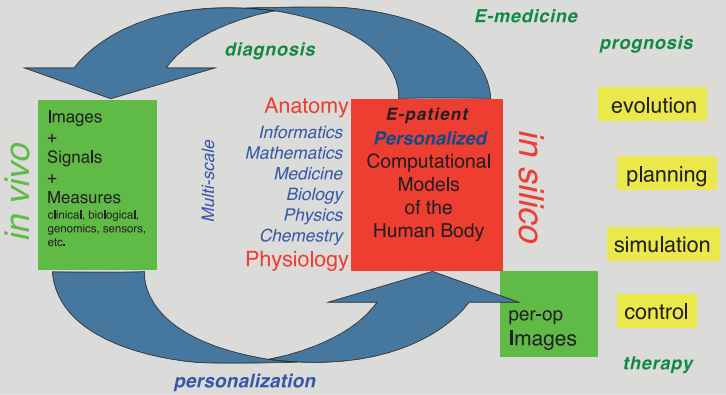
1. Análisis de imágenes biomédicas y aprendizaje automático

2. Imágenes y fenómica, bioestadística

3. Anatomía computacional, estadística geométrica

4. Fisiología computacional y terapia guiada por imágenes

5. Cardiología computacional e intervenciones cardíacas basadas en imágenes



DANTE

Redes dinámicas: Enfoque de captura temporal y estructural

El equipo de DANTE desarrolla técnicas de aprendizaje automático y algoritmos de procesamiento de señales con el objetivo principal de

dotarlos de sólidos fundamentos teóricos, interpretabilidad física y eficiencia de recursos.

Con una cultura arraigada en la interfaz del procesamiento de señales y el aprendizaje automático, la experiencia del equipo aprovecha la noción de parsimonia y sus variantes estructuradas –y notablemente la de los grafos– que desempeñan un papel fundamental para garantizar la identificabilidad de las descomposiciones en espacios latentes, como los problemas inversos en el procesamiento de señales de alta dimensión.

Los logros recientes del equipo incluyen algoritmos distribuidos para aprender de representaciones de datos altamente comprimidas con garantías de privacidad, y técnicas para explotar paseos aleatorios en grafos para el aprendizaje semi-supervisado en entornos difíciles. Uno de los principales retos es aprovechar estas ideas para garantizar no sólo métodos eficientes en cuanto a recursos, sino también decisiones explicables y parámetros aprendidos interpretables, todos los cuales son grandes desafíos sociales para que las "decisiones algorítmicas" sean fiables y aceptables.

Los retos del **análisis de señales para el habla y el sonido** tienen mucho en común con aquellos indicados anteriormente: la escalabilidad, la multimodalidad y la introducción de conocimientos previos son relevantes también para las aplicaciones de audio. Las aplicaciones objetivo son la identificación de hablantes, la comprensión del habla, el diálogo –incluso para robots–, la separación de fuentes (en el caso de conversaciones múltiples), el reconocimiento y la síntesis de emociones y la traducción automática en tiempo real. En el caso de las señales de audio, también es obligado desarrollar o tener acceso a un gran volumen de datos para el aprendizaje automático. El aprendizaje incremental en línea podría ser necesario para el procesamiento del habla en tiempo real.

PERCEPTION

Interpretación y modelización de imágenes y sonidos

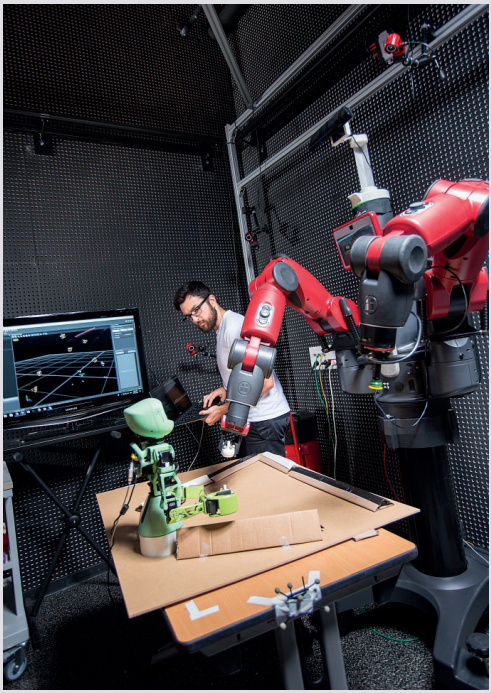
El programa de investigación del grupo PERCEPTION consiste en la investigación y aplicación de modelos informáticos para la asignación de imágenes y sonidos al significado y a las acciones. Los miembros

del equipo PERCEPTION abordan este desafiante problema con un enfoque interdisciplinario que abarca los siguientes temas: visión por computador, procesamiento de señales auditivas, análisis de escenas de audio, aprendizaje automático y la robótica. En concreto, desarrollamos métodos para la representación y el reconocimiento de objetos y eventos visuales y auditivos, la fusión audiovisual, el reconocimiento de las acciones humanas, los gestos y el habla, la audición espacial y la interacción entre humanos y robots.

Temas de investigación:

- **Visión por computador:** representación espacio-temporal de información visual en 2D y 3D, reconocimiento de acciones y gestos, análisis de rostros humanos, sensores 3D, visión binocular, sistemas de cámaras múltiples, seguimiento de personas y objetos en secuencias de vídeo.

- **Análisis de escenas auditivas:** audición binocular, localización de múltiples fuentes de sonido, seguimiento y separación, comunicación del habla, clasificación de eventos sonoros, diarización de hablantes, mejora de la señal acústica.



Poppy Torso aprendiendo a hablar con Baxter Mommy - © Inria _ Photo C. Morel

- **Aprendizaje automático:** modelos mixtos probabilísticos, reducción de dimensión lineal y no lineal, aprendizaje combinado, modelos gráficos, inferencia bayesiana, redes neuronales y aprendizaje profundo.

- **Robótica:** visión de los robots, audición de los robots, interacción entre robots y humanos, fusión de datos, arquitecturas de software.

Los retos específicos en el ámbito del habla son:

Uso de modelos pre-entrenados que se auto-supervisen para el reconocimiento del habla

La aplicación de métodos de pre-entrenamiento auto-supervisado al habla podría brindar en los próximos años resultados tan espectaculares como en el caso de textos, con muchas aplicaciones en el campo del procesamiento automático del habla en lenguas de escasos recursos (algunas de las cuales carecen de recursos textuales). En general, la aplicación del aprendizaje automático a las lenguas o culturas económicamente no dominantes es muy importante para evitar que se amplíe la brecha digital.

Procesar señales de audio del "mundo real"

El procesamiento automático de la señal de audio real es un problema no resuelto (al contrario de lo que podría pensarse). La separación de fuentes no funciona bien "en la naturaleza". En consecuencia, la caída del rendimiento del procesamiento automático del lenguaje junto con sonidos ambientales no permite toda una serie de aplicaciones médicas o educativas. Por lo general, el aprendizaje automático debe aprender a salir del marco de los datos encajonados y enfrentarse de cara al difícil problema de los datos reales si se quiere utilizar en aplicaciones concretas.

MULTISPEECH

Modelado del habla para facilitar la comunicación oral

Más allá del aprendizaje supervisado de caja negra – MULTISPEECH estudia retos fundamentales relacionados con el aprendizaje profundo. Por ejemplo, exploran métodos híbridos que combinan el aprendizaje profundo con el modelado estadístico, el procesamiento de señales o el razonamiento simbólico para aumentar el rendimiento y la explicabilidad, diseñan métodos de aprendizaje débilmente supervisado o de aprendizaje de transferencia para explotar etiquetas mal clasificadas (noisy) o datos fuera del dominio, y exploran métodos de anonimización del habla para preservar la privacidad de los sujetos de los datos.

Producción del habla – MULTISPEECH desarrolla un sistema de síntesis articulatoria del habla basado en el modelado de la dinámica del tracto vocal, y una cabeza parlante muy realista basada en la animación dinámica de la boca y las expresiones faciales. Entre sus aplicaciones se encuentran la animación por computador y el aprendizaje del lenguaje para niños con dificultades o hipoacúsicos.

El habla en su entorno – MULTISPEECH diseña algoritmos para mejorar el habla en presencia de un eco acústico, reverberación, ruido y hablantes en competencia, y para lograr un reconocimiento robusto del habla y del hablante en tales condiciones. Modela la semántica para mejorar aún más el reconocimiento y clasificar los contenidos hablados. Por último, desarrolla métodos para estimar las propiedades acústicas de la sala y detectar eventos sonoros del entorno. Más allá de la comunicación hablada, estos métodos tienen muchas aplicaciones en cuanto a la monitorización del sonido, la audición de robots, la acústica de edificios, la realidad aumentada o la monitorización de medios sociales.

PANAMA

Parsimonia y nuevos algoritmos para el modelado de señales y audio

En la interfaz entre el modelado de audio y el procesamiento matemático de la señal, el objetivo global de PANAMA es desarrollar soluciones matemáticamente fundadas y algorítmicamente eficientes para modelar, adquirir y procesar señales de alta dimensión, con un fuerte énfasis en los datos acústicos.

Las aplicaciones alimentan los marcos matemáticos y estadísticos propuestos con escenarios prácticos, y se realizan pruebas extensas de los algoritmos desarrollados en aplicaciones específicas. La metodología de PANAMA se basa en un bucle cerrado entre las investigaciones teóricas, el desarrollo de algoritmos y los estudios empíricos.

Los fundamentos científicos de PANAMA se centran en las

representaciones dispersas y el modelado probabilístico, y su alcance científico se orienta hacia tres grandes direcciones:

- La ampliación del paradigma de la representación dispersa hacia el del "modelado disperso", con el reto de establecer, reforzar y aclarar las conexiones entre las representaciones dispersas y el aprendizaje automático.

- Un enfoque en modelos probabilísticos sofisticados y métodos estadísticos avanzados para dar cuenta de las dependencias complejas entre las variables de múltiples capas (como en los flujos audiovisuales, contenidos musicales, datos biomédicos, la teledetección, etc.).

- La investigación de las representaciones, procesamiento y transformaciones basadas en grafos, con el objetivo de describir, modelar e inferir las estructuras subyacentes dentro de los flujos de contenidos o conjuntos de datos.

Acciones exploratorias (AExs)

AEx- Ayana – IA y teledetección a bordo para el Nuevo Espacio

El AEx de AYANA es un proyecto interdisciplinar que utiliza conocimientos del modelado estocástico, procesamiento de imágenes, inteligencia artificial, teledetección y electrónica/computación integrada. La industria aeroespacial está en plena expansión y enfrentando cambios (el "Nuevo Espacio"). Actualmente está experimentando numerosos cambios en relación con los sensores a nivel espectral (IRT no refrigerado, ultravioleta lejano, etc.) y a nivel material (la llegada de las nanotecnologías o la nueva generación de "Systems on Chips" (SoCs) por ejemplo), como con los portadores de estos sensores: satélites geoestacionarios de alta resolución; satélites de baja órbita tipo Leo; o minisatélites y cubosatélites industriales en conjunción. AYANA manejará un gran número de datos, los que consisten en imágenes de gran tamaño, con resoluciones y componentes espectrales muy variados y que forman series temporales a frecuencias de 1 a 60 Hz. En cuanto al aspecto de la electrónica/computación embebida, AYANA trabajará en estrecha colaboración con especialistas en la materia ubicados en Europa, que trabajan en agencias espaciales y/o para contratistas industriales.

AEx- ACOUST.IA – Inteligencia Artificial para apoyar la acústica de edificios

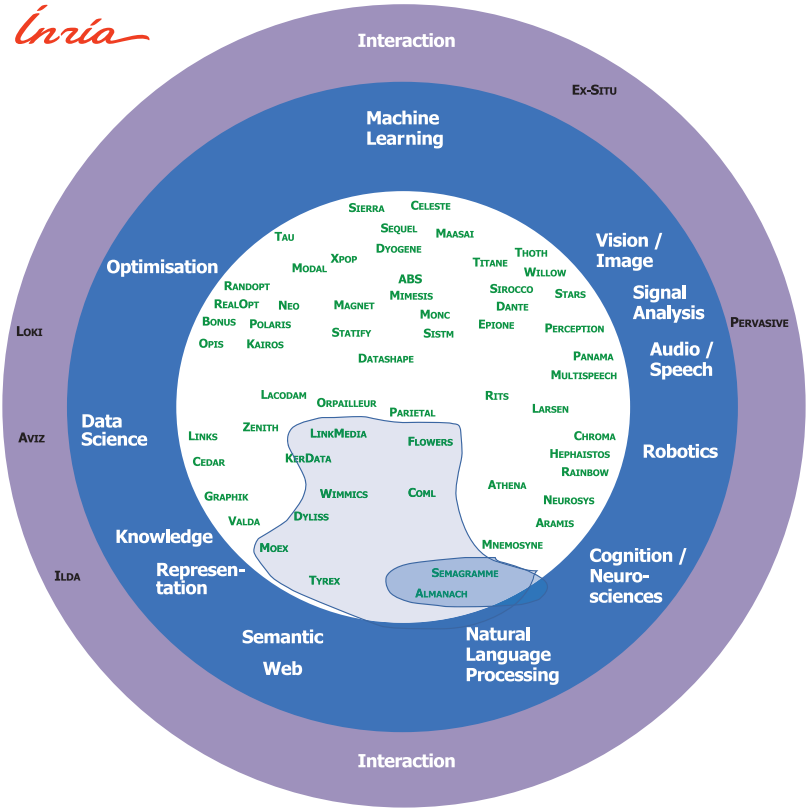
Equipo-proyecto: MULTISPEECH

¿Es posible establecer el perfil acústico de una sala con la simple grabación de un aplauso? Este es el objetivo de ACOUST.IA, que pretende simplificar y mejorar radicalmente la precisión del diagnóstico acústico de los edificios, un importante problema de salud pública, gracias a la inteligencia artificial y el procesamiento de señales. Se desarrollarán enfoques innovadores que combinen el aprendizaje supervisado, el modelado estadístico y físico y el procesamiento de audio multicanal para superar las limitaciones de los enfoques manuales, costosos e iterativos que se utilizan actualmente.

Otros equipos-proyectos en este ámbito: TITANE (Sophia Antipolis), MORPHEO (Grenoble)

5.4 Procesamiento del lenguaje natural

El campo del Procesamiento del Lenguaje Natural (NLP, por sus siglas en inglés) se remonta a los años cincuenta. No obstante, sigue teniendo una relevancia esencial para la nueva sociedad de la información. Su objetivo es procesar textos en lenguaje natural, ya sea para analizar textos existentes o generar otros nuevos, o para lograr un procesamiento del lenguaje similar al del humano para una serie de tareas o aplicaciones. Estas aplicaciones, reagrupadas bajo el término "ingeniería del lenguaje", incluyen la traducción automática, la respuesta a preguntas, recuperación de información, extracción de información, minería de textos, apoyo en la lectura y la escritura y muchas otras. Desde un punto de vista más orientado a la investigación, la lingüística empírica y las humanidades digitales también pueden considerarse ámbitos de aplicación del NLP.



El NLP es un ámbito interdisciplinario; requiere conocimientos de lingüística formal y descriptiva (para desarrollar modelos lingüísticos de las lenguas humanas), de ciencias de la computación y algoritmos (para diseñar y desarrollar programas eficientes que puedan tratar dichos modelos) y de matemáticas aplicadas (para adquirir automáticamente conocimientos lingüísticos o generales). El procesamiento de textos en lenguaje natural es una tarea difícil, sobre todo por la gran cantidad de ambigüedad que existe en éste, las especificidades de cada lengua y dialecto y porque muchos usuarios no se ajustan necesariamente a las convenciones gramaticales y ortográficas, cuando éstas existen.

Las primeras décadas de la implementación del NLP se centraron sobre todo en los enfoques simbólicos, con aportes también de nociones importantes a la informática, especialmente en relación con la teoría de la gramática formal y las técnicas de análisis sintáctico. El conocimiento lingüístico se codificaba principalmente en forma de gramáticas y bases de datos léxicas desarrolladas manualmente. En las dos últimas décadas, los enfoques estadísticos y aquellos basados en el aprendizaje automático (word embedding, RNN, transformers) han renovado en gran medida el campo, empujando los corpus anotados para que ocupen el centro de escena y mejorando significativamente el estado del arte.

Hibridación entre el aprendizaje automático (ML) y los modelos simbólicos

A pesar de los importantes avances realizados en los últimos años, los estudios del diálogo natural siguen dando resultados poco impresionantes. Adolecen de muchos problemas (por ejemplo, un problema mal planteado, falta de métricas de evaluación y dificultad para generalizar fuera del conjunto de entrenamiento). Sin embargo, uno de los problemas centrales es también el de considerar el diálogo como un problema de aprendizaje automático puro, en cuanto que colocar al ser humano dentro de este bucle es esencial, lo que implica dialogar con otras disciplinas (ciencias sociales, ciencias cognitivas, etc.). Los enfoques simbólicos conservan ventajas específicas y los mejores resultados podrían obtenerse al aprovechar todo tipo de recursos dentro de sistemas híbridos que acoplen técnicas simbólicas y estadísticas.

ALMANACH

Modelado y análisis automático del lenguaje y humanidades computacionales

El equipo-proyecto ALMAnaCH (ALMAnaCH se creó como equipo Inria ("équipe") el 1 de enero de 2017 y como equipo-proyecto el 1 de julio de 2019) reúne a especialistas de un ámbito de investigación pluridisciplinario en la interfaz entre la informática, la lingüística, la estadística y las humanidades, a saber, el **procesamiento del lenguaje natural, la lingüística computacional y las humanidades y ciencias sociales digitales y computacionales.**

La lingüística computacional es un campo interdisciplinario que se ocupa de la modelización computacional del lenguaje natural. La

investigación en este campo está motivada tanto por el objetivo teórico de comprender el lenguaje humano como por las aplicaciones prácticas en el **Procesamiento del Lenguaje Natural** (NLP), como el análisis lingüístico (análisis sintáctico y semántico, por ejemplo), la traducción automática, la extracción y recuperación de información y el diálogo entre personas y computadores. La lingüística computacional y el NLP, que se remontan al menos a principios de los años 50, se encuentran entre los sub-campos clave de la **Inteligencia Artificial.**

Las Humanidades Digitales y Ciencias Sociales (DH por sus siglas en inglés) es un campo interdisciplinario que utiliza la informática como fuente de técnicas y tecnologías, en particular el NLP, para explorar cuestiones de investigación en las ciencias sociales y las humanidades. Las **Humanidades Computacionales** y las ciencias sociales computacionales tienen como objetivo mejorar el estado del arte tanto en las ciencias de la computación (por ejemplo, el NLP) como en las ciencias sociales y las humanidades, mediante la participación de la informática como campo de investigación.

Uno de los principales retos de la lingüística computacional es **modelar y hacer frente a la variación del lenguaje**. El lenguaje varía en función del ámbito y el género (noticias, literatura científica, poesía, transcripciones orales, etc.), de factores sociolingüísticos (edad, procedencia, educación; variante atestiguada, por ejemplo, en las redes sociales), factores geográficos (dialectos) y otras dimensiones (discapacidades, por ejemplo). Sin embargo, el lenguaje también evoluciona constantemente en todas las escalas. Abordar esta variabilidad sigue siendo una cuestión pendiente para el NLP. Los enfoques habituales, que suelen basarse en métodos de aprendizaje automático supervisado y semi-supervisado, requieren grandes cantidades de datos anotados. Siguen adoleciendo del alto nivel de variabilidad que se encuentra, por ejemplo, en los **contenidos generados por usuarios**, en los textos **no contemporáneos** y en los **documentos de dominios específicos** (por ejemplo, financieros o jurídicos).

SEMAGRAMME

Análisis semántico del lenguaje natural

La lingüística computacional es una disciplina que se encuentra en la intersección de la informática y la lingüística. En su vertiente teórica, pretende ofrecer modelos computacionales de la facultad del lenguaje humano. En relación con su aplicación, se ocupa del procesamiento del lenguaje natural y sus usos prácticos.

El programa de investigación de Sémagramme pretende desarrollar modelos basados en matemáticas bien establecidas. Buscamos dos ventajas principales en este enfoque. Por una parte, al basarnos en teorías maduras, tenemos a nuestra disposición conjuntos de herramientas matemáticas que podemos utilizar para estudiar nuestros modelos. Por otra parte, desarrollar varios modelos sobre un fondo matemático común facilitará su integración y la búsqueda de principios unificadores.

Los principales dominios matemáticos en los que nos basamos son la teoría de los lenguajes formales, la lógica simbólica y la teoría de tipos.

La teoría de los lenguajes formales estudia los aspectos puramente sintácticos y combinatorios de los lenguajes, vistos como conjuntos de cadenas (o posiblemente árboles o grafos). La teoría de los lenguajes formales ha sido especialmente fructífera para el desarrollo de algoritmos de análisis sintáctico de lenguajes libres de contexto. Nosotros la utilizamos, de forma similar, para desarrollar algoritmos de análisis sintáctico para formalismos que van más allá de la ausencia de contexto. La teoría de los lenguajes formales también parece ser muy útil para estudiar formalmente la capacidad expresiva y la complejidad de los modelos que desarrollamos.

La lógica simbólica (y, más concretamente, la teoría de la demostración) se ocupa del estudio del poder expresivo y deductivo de los sistemas formales. En un enfoque de la lingüística computacional basado en reglas; el uso de la lógica simbólica es omnipresente. Como hemos dicho anteriormente, a nivel de la sintaxis, se pueden considerar varios tipos de gramáticas (generativas o categoriales) como sistemas deductivos

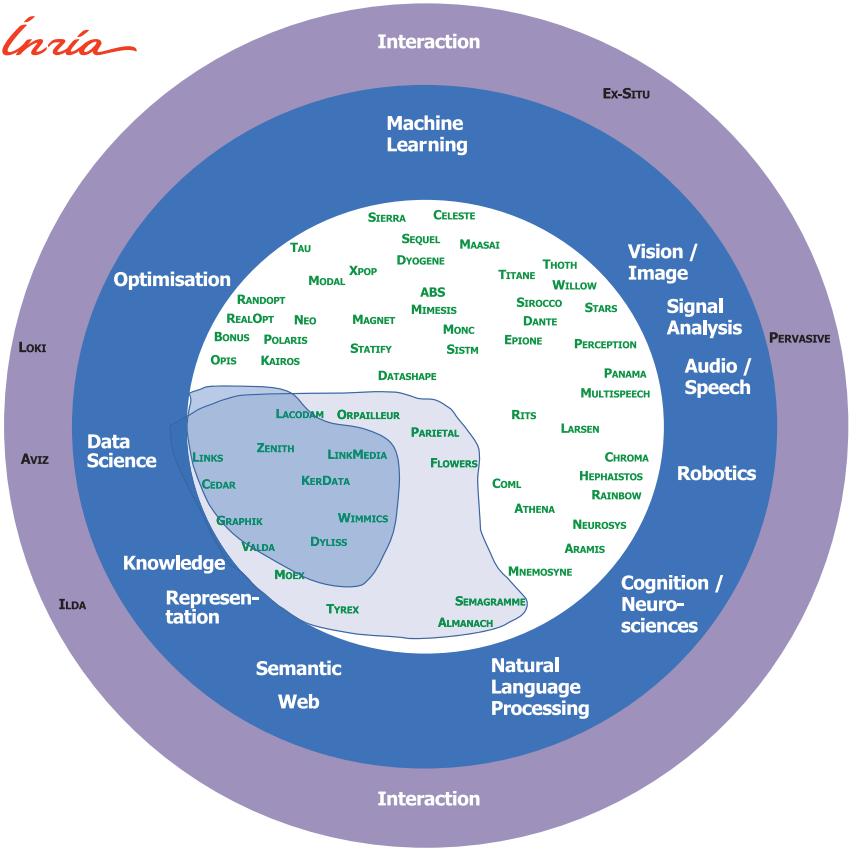
básicos. Al nivel de la semántica, el significado de un enunciado se capta computando representaciones semánticas (intermedias) que se expresan como formas lógicas. Por último, el uso de la lógica simbólica permite formalizar las nociones de inferencia y vinculación que se necesitan a nivel de la pragmática.

Entre las diversas lógicas posibles que se pueden utilizar, el cálculo lambda (λ-cálculo) de tipado simple de Church y la teoría simple de tipos (también conocida como lógica de orden superior) desempeñan un papel central. Por un lado, la semántica de Montague se basa en el λ-cálculo de tipado simple, como también lo hace nuestro modelo de interfaz sintaxis-semántica. Por otro lado, como ha mostrado Gallin, la lógica objetivo utilizada por Montague para expresar significados (es decir, su lógica intensional) es esencialmente una variante de la lógica de orden superior que presenta tres tipos atómicos (el tercero de éstos representa el conjunto de mundos posibles).

5.5 Sistemas basados en el conocimiento y web semántica

Según la definición inicial de Tim Berners-Lee, *"la Web Semántica es una extensión de la web actual en la que la información recibe un significado bien definido, lo que permite a los computadores y a las personas trabajar en cooperación."* La torre semántica se basa en URIs y XML, mediante esquemas RDF que representan tripletas de datos, hasta llegar a las ontologías que permiten el razonamiento y el tratamiento lógico.

Los equipos de Inria dedicados a la representación, el razonamiento y procesamiento del conocimiento abordan los siguientes retos de diferentes maneras: (i) tratar grandes volúmenes de información procedentes de fuentes distribuidas heterogéneas; (ii) construir puentes entre datos masivos almacenados en bases de datos utilizando tecnologías semánticas; y (iii) desarrollar aplicaciones basadas en la semántica sobre estas tecnologías.



El manejo de grandes volúmenes de información procedentes de fuentes distribuidas heterogéneamente

Gracias a la omnipresencia del Internet, nos enfrentamos a la oportunidad y el reto de cambiar de sistemas inteligentes artificiales locales a inteligencias y sociedades artificiales ampliamente diseminadas. El diseño y funcionamiento de sistemas fiables y eficientes que combinen datos enlazados procedentes de fuentes distantes mediante flujos de trabajo de servicios descentralizados sigue siendo un problema no resuelto. La calidad de los datos y la trazabilidad de sus procesos, la precisión de su extracción y captura, la corrección de su alineación e integración, la disponibilidad y calidad de los modelos compartidos (ontologías, vocabularios) para representarlos, intercambiarlos y razonar sobre ellos, etc., son aspectos que deben abordarse a gran escala y de forma continua.

Un segundo aspecto respecta a la Web, que no sólo proporciona un marco de aplicación universal para el Internet, sino también un espacio híbrido en el que los humanos y agentes de software pueden interactuar a gran escala y formar comunidades mixtas. En la actualidad, millones de usuarios y agentes artificiales interactúan a diario en aplicaciones en línea, lo que da lugar a sistemas muy complejos que hay que estudiar y diseñar. Necesitamos modelos y algoritmos que generen justificaciones y explicaciones y acepten retroalimentación para apoyar las interacciones con usuarios muy diferentes. Tenemos que considerar los sistemas complejos incluyendo a los usuarios como un componente inteligente que interactuará con otros componentes (por ejemplo, la inteligencia artificial en las interfaces, la interacción del lenguaje natural), participará en el proceso (por ejemplo, la computación humana, crowdsourcing, las máquinas sociales) y que puede ser aumentado por el sistema (amplificación de la inteligencia, aumento cognitivo, inteligencia aumentada, mente extendida y cognición distribuida) .

WIMMICS

Interacciones hombre-máquina, comunidades y semántica instrumentadas en la Web

La Web ofrece espacios virtuales (por ejemplo, Wikipedia) en los que las personas y los programas informáticos interactúan en comunidades mixtas que intercambian y utilizan conocimientos formales (por ejemplo, ontologías, bases de conocimiento) y contenidos informales (por ejemplo, textos, mensajes, etiquetas).

El equipo de WIMMICS estudia modelos y métodos para unir la semántica formal con la semántica social en la web. Sigue un enfoque multidisciplinario para analizar y modelar estos espacios, sus comunidades de usuarios y sus interacciones. También proporciona algoritmos para calcular estos modelos a partir de rastros en la web, incluyendo la extracción de conocimiento del texto, el análisis semántico de redes sociales y la teoría de la argumentación.

Para formalizar y razonar en base a estos modelos, el equipo de WIMMICS propone lenguajes y algoritmos que se basan en enfoques de conocimiento basados en grafos para la web semántica y los datos enlazados en la Web, por ejemplo, los modelos de grafos del Marco de Descripción de Recursos (RDF, por sus siglas en inglés). En conjunto,

estas contribuciones proporcionan herramientas de análisis e indicadores, y apoyan nuevas funcionalidades y tareas de gestión en las comunidades epistémicas.

Se puede agrupar los objetivos de investigación de WIMMICS según cuatro temas que identificamos en la conciliación de la semántica social y formal en la web:

Tema 1 – Modelización de los usuarios y diseño de la interacción en la Web: La pregunta general de investigación que aborda este objetivo es ¿cómo podemos mejorar nuestras interacciones con una web semántica y social cada vez más compleja y densa? WIMMICS se centra en preguntas subordinadas específicas: ¿Cómo podemos capturar y modelar las características de los usuarios? ¿Cómo podemos representar y razonar con los perfiles de los usuarios? ¿Cómo podemos adaptar los comportamientos del sistema como resultado de lo anterior? ¿Cómo podemos diseñar nuevos medios de interacción? ¿Cómo podemos evaluar la calidad de la interacción diseñada?

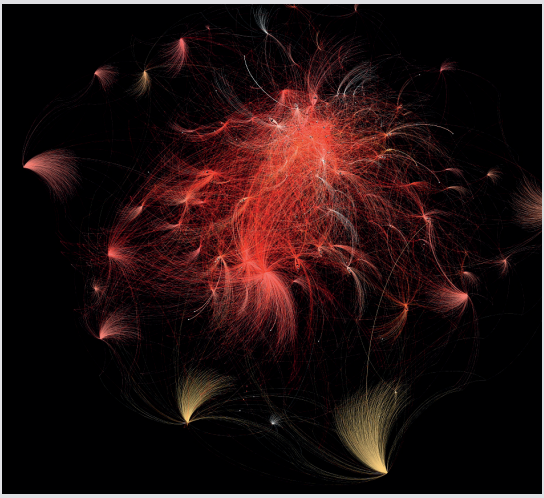
Tema 2 – Análisis de las comunidades e interacciones sociales en la Web: La pregunta general que se aborda en este segundo objetivo es ¿Cómo podemos gestionar la actividad colectiva en los medios sociales? WIMMICS se centra en las siguientes preguntas subordinadas: ¿Cómo analizamos las prácticas de interacción social y las estructuras en las que estas prácticas tienen lugar? ¿Cómo captamos las interacciones y estructuras sociales? ¿Cómo podemos formalizar los modelos de estas construcciones sociales? ¿Cómo podemos analizar y razonar sobre estos modelos de la actividad social?

Tema 3 – Vocabularios, representación del conocimiento basada en la web semántica y los datos enlazados y formalismos de la Inteligencia Artificial en la web: La pregunta general que se aborda en este tercer objetivo es ¿cuáles son los esquemas y extensiones necesarios de los formalismos de la web semántica para nuestros modelos? WIMMICS se centra en varias preguntas subordinadas: "¿Qué clase de formalismos son los más adecuados para los modelos de la sección anterior? ¿Cuáles son las limitaciones y posibles extensiones de los formalismos existentes? ¿Cuáles son los esquemas, ontologías y vocabularios que faltan? ¿Cuáles son los vínculos y las posibles combinaciones entre los

formalismos existentes? En pocas palabras, una parte importante de este objetivo es formalizar como grafos tipificados los modelos identificados en los objetivos anteriores para que el software los explote en su procesamiento (en el siguiente objetivo).

Tema 4 – Procesamiento de la inteligencia artificial: aprendizaje, análisis y razonamiento sobre grafos semánticos heterogéneos en la Web: La pregunta general de investigación que se aborda en este último objetivo es ¿cuáles son los algoritmos necesarios para analizar y razonar sobre los grafos heterogéneos que obtenemos? WIMMICS se centra en varias preguntas subordinadas: "¿Cómo analizamos grafos de diferentes tipos y sus interacciones? ¿Cómo soportamos los diferentes ciclos de vida de los grafos, los cálculos y las características de una manera coherente y comprensible? ¿Qué tipo de algoritmos pueden apoyar las diferentes tareas de nuestros usuarios?

Los resultados de estas investigaciones se integran, evalúan y transfieren a través de software genérico (por ejemplo, la fábrica de web semántica CORESE) y aplicaciones específicas (por ejemplo, CREEP para detectar el ciberacoso). El objetivo final del equipo es hacer de la web un lugar en el que se puede vincular a la perfección la inteligencia natural con la artificial.



Grafo de datos del motor de búsqueda exploratoria Discovery Hub - © Inria _ WIMMICS

En efecto, los datos producidos y los conocimientos extraídos cambian constantemente, por lo que los agentes y procesos que los consumen deben ser capaces de adaptar sus propios conocimientos.

MOEX

Conocimiento en evolución

MOEX estudia los principios mediante los que evoluciona el conocimiento de los agentes sociales. Estos agentes pueden ser programas que observan la web (semántica), seleccionan e intercambian información interesante o robots sociales que se comunican con humanos y otros robots. Toi.Net parece cubrir ambos casos. Los agentes se enfrentan a entornos cambiantes (a Sam ya no le importa el concurso de Eurovisión, nuevos conocimientos sobre los coronavirus) y pueden tener que interactuar con otros agentes (Sam, nuevos amigos de Sam u otros robots).

El comportamiento de estos agentes se rige por conocimientos que pueden representarse de diversas maneras. En una situación cambiante, los agentes no deben esperar a que un programador actualice sus conocimientos o a que se generen muchos ejemplos y se cometan otros tantos errores. Deben adaptar sus conocimientos para comportarse adecuadamente. Los mecanismos de adaptación del conocimiento responden a la presión externa, ejercida por el entorno y la sociedad en la que evolucionan los agentes y a la presión interna para garantizar la coherencia del conocimiento.

El objetivo es responder, en particular, a las siguientes preguntas:

- ¿Cómo adaptan las poblaciones de agentes su representación del conocimiento a su entorno y a otras poblaciones?

- ¿Cómo debe evolucionar este conocimiento cuando el entorno cambia y se encuentran nuevas poblaciones?

- ¿Cómo pueden los agentes preservar la diversidad de conocimientos y es esta diversidad beneficiosa?

Para ello, combinamos los métodos de representación del conocimiento y de evolución cultural. Los primeros proporcionan modelos formales de conocimiento; los segundos, un marco bien definido para estudiar la evolución situada. Consideramos que el conocimiento es una cultura y estudiamos las propiedades globales de los operadores de adaptación local aplicados por poblaciones de agentes al realizar lo siguiente de forma conjunta:

- Comprobar experimentalmente las propiedades de los operadores de adaptación en diversas situaciones mediante la evolución cultural experimental, y

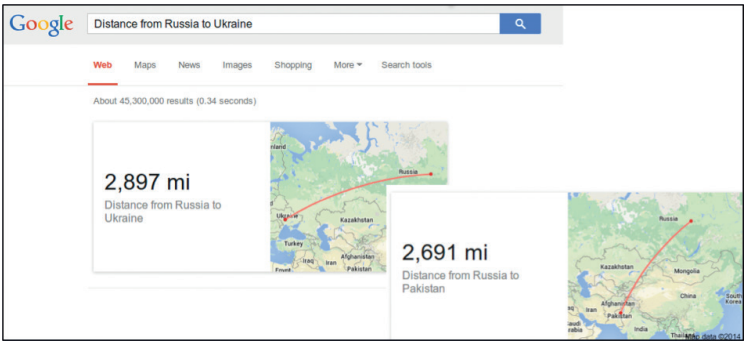
- Determinar teóricamente dichas propiedades modelando cómo los operadores dan forma a la representación del conocimiento.

Nuestro objetivo es adquirir una comprensión precisa de la evolución del conocimiento mediante la consideración de una amplia gama de situaciones, representaciones y operadores de adaptación

Construir puentes entre datos masivos almacenados en bases de datos utilizando tecnologías semánticas

La web semántica aborda la integración masiva de fuentes de datos muy distintas (por ejemplo, sensores de ciudades inteligentes, conocimiento biológico extraído de artículos científicos, descripciones de eventos en redes sociales) y utilizando vocabularios muy diferentes (por ejemplo, esquemas relacionales, tesauros ligeros, ontologías formales) en razonamientos muy diferentes (por ejemplo, toma de decisiones por derivación lógica, enriquecimiento por inducción, análisis mediante minería, etc.). En la Web, al grafo inicial de páginas enlazadas se ha adherido un número creciente de otros grafos y ahora se mezclan con sociogramas que capturan la estructura de la red social, flujos de trabajo que especifican las vías de decisión a seguir, registros de búsqueda que capturan los rastros de nuestra navegación, composiciones de servicios que especifican el procesamiento distribuido, datos abiertos que enlazan conjuntos de datos distantes, etc. Además, estos grafos no están disponibles en un único repositorio central, sino que están distribuidos en muchas fuentes diferentes y algunos subgrafos son públicos (por ejemplo, dbpedia http://dbpedia.org) mientras que otros son privados (por ejemplo, datos corporativos). Algunos subgrafos son pequeños y

locales (por ejemplo, el perfil de un usuario en un dispositivo), otros son enormes y están alojados en clusters o agrupaciones (por ejemplo, Wikipedia), otros son muy estables (por ejemplo, el tesauro del latín), otros cambian varias veces por segundo (por ejemplo, los estados de las redes sociales), etc. Cada tipo de red de la web no es una isla aislada, sino que interactúan entre sí: las redes sociales influyen en los flujos de mensajes, sus temas y tipos, los enlaces semánticos entre términos interactúan con los enlaces entre sitios y viceversa, etc. El reto es enorme, no sólo para encontrar los medios de representar y analizar cada tipo de grafos, sino también para combinarlos y combinar su tratamiento.



Extraído del artículo "Why the Data Train Needs Semantic Rails" de Janowicz et al., AI Magazine, 2015. Sin la semántica, Rusia parece más cercana a Pakistán que a Ucrania.

CEDAR

Exploración de datos enriquecidos a escala de la nube

Para dar sentido al "Big Data", es necesario interpretarlo a través del prisma del conocimiento en cuanto al contenido, la organización y el significado de los datos. Además, el conocimiento del dominio es a menudo el lenguaje más cercano a los usuarios, ya sean éstos expertos del dominio o usuarios finales novatos de una aplicación de datos intensivos. Por ello, las herramientas expresivas y escalables para OBDA (Ontology-Based Data Access o acceso a datos basado en ontologías) son un factor clave para el éxito de las aplicaciones de Big Data.

Cedar trabaja en la interfaz entre los formalismos de representación del conocimiento (como algunas lógicas de descripción o clases de reglas

existenciales) y los motores de bases de datos. El equipo construye herramientas OBDA altamente eficientes con un enfoque particular en el escalamiento de bases de datos muy grandes; esto puede ser visto como el aumento de los motores de bases de datos con capacidades de razonamiento, y su despliegue en un entorno de nube para la escala. Cedar también investiga nuevas formas de interactuar con bases de datos y conocimientos grandes y complejos, como los referenciados en la nube de Linked Open Data (http://lod-cloud.net). También se investiga la semántica como medio para integrar y dar sentido a contenidos heterogéneos y complejos, en repositorios de datos web ricos y heterogéneos, en particular aplicados a la comprobación de hechos en el área del periodismo.

Optimización y rendimiento a escala: este tema yace en el centro del proyecto del ERC (Consejo Europeo de Investigación) de Yanlei Diao "Big and Fast Data", cuyo objetivo es la optimización con garantías de rendimiento para el procesamiento de datos en tiempo real en la nube. Las técnicas de aprendizaje automático y la optimización multiobjetivo se aprovechan para construir modelos de rendimiento para el análisis de datos en la nube. El mismo objetivo es compartido por nuestro trabajo sobre la evaluación eficiente de consultas en bases de conocimiento dinámicas.

El descubrimiento y exploración de datos: el Big Data actual es complejo; entenderlo y explotarlo es difícil. Para ayudar a los usuarios, exploramos resúmenes compactos de bases de conocimiento para abstraer su estructura y ayudar a los usuarios a formular consultas; realizamos la exploración interactiva de grandes bases de datos relacionales; exploramos técnicas para descubrir información interesante en las bases de conocimiento; y estudiamos técnicas de la búsqueda de palabras clave sobre fuentes de Big Data.

Minería de grafos de datos
© Inria _ Photo S. Erôme - Signatures

Graphik

Grafos para inferencia y representación del conocimiento

El principal ámbito de investigación de GraphIK es la Representación del Conocimiento y el Razonamiento (KR por sus siglas en inglés), que estudia los paradigmas y formalismos para representar el conocimiento y razonar sobre estas representaciones. Gran parte de nuestro trabajo está estrechamente relacionado con la gestión de datos y la teoría de bases de datos.

Desarrollamos lenguajes lógicos, que corresponden principalmente a fragmentos de la lógica de primer orden. Sin embargo, también utilizamos grafos e hipergrafos (en el sentido de la teoría de grafos) como objetos básicos. Efectivamente, vemos los grafos etiquetados como una representación abstracta del conocimiento que puede expresarse en muchos lenguajes de KR: diferentes tipos de grafos conceptuales– históricamente nuestro principal objetivo– el lenguaje de la Web Semántica RDFS, reglas expresivas equivalentes a las llamadas dependencias generadoras de tuplas en las bases de datos, algunas lógicas descriptivas dedicadas a la respuesta de consultas, etc. Para estos lenguajes, el razonamiento puede basarse en la estructura de los objetos (por tanto, en nociones grafo-teóricas), siendo al mismo tiempo sólido y completo con respecto a la vinculación en los fragmentos lógicos asociados. Una cuestión importante es estudiar las compensaciones entre la expresividad y la trazabilidad computacional del razonamiento (sólido y completo) en estos lenguajes.

GraphIK se centra en algunos de los principales retos del KR:

- Respuesta a consultas ontológicas; consulta de conjuntos de datos grandes, complejos o heterogéneos, dotados de una capa ontológica;
- Razonamiento con lenguajes basados en reglas;
- Razonamiento en presencia de incoherencias; y
- La toma de decisiones.

Una característica importante de las técnicas basadas en el conocimiento es su poder explicativo, es decir, su capacidad potencial para explicar las conclusiones extraídas. Ser capaz de explicar, justificar o argumentar es un requisito obligatorio en muchas aplicaciones de IA en las que los usuarios necesitan entender los resultados del sistema, para poder confiar en él y controlarlo. Además, se convierte en una preocupación esencial con respecto a las cuestiones éticas en cuanto las decisiones automatizadas pueden afectar a los seres humanos.

LINKS

Enlazando datos dinámicos

La aparición de datos enlazados en la web exige nuevas tecnologías de gestión de bases de datos para las colecciones de datos enlazados. Los retos clásicos de la investigación en bases de datos deben plantearse ahora para los datos enlazados: cómo definir consultas lógicas exactas, cómo gestionar las actualizaciones dinámicas y cómo automatizar la búsqueda de consultas adecuadas. A diferencia de la corriente principal de datos abiertos enlazados, el proyecto LINKS se centra en colecciones de datos enlazados en varios formatos, arraigado al supuesto de que los datos son correctos en la mayoría de las dimensiones. Los retos siguen siendo difíciles como consecuencia de los datos incompletos, a los esquemas poco informativos o heterogéneos y a los errores y ambigüedades que siguen existiendo en los datos. Desarrollamos algoritmos para evaluar y optimizar consultas lógicas sobre colecciones de datos enlazados, algoritmos incrementales que pueden monitorear flujos de datos enlazados y gestionar actualizaciones dinámicas de colecciones de datos enlazados, y algoritmos de aprendizaje simbólico que pueden deducir consultas apropiadas para colecciones de datos enlazados a partir de ejemplos.

Temas de investigación

Desarrollamos algoritmos para responder a consultas lógicas sobre colecciones de datos enlazados heterogéneos en formatos híbridos, lenguajes de programación distribuidos para gestionar colecciones de

datos enlazados dinámicos y flujos de trabajo basados en consultas y mapeos, y algoritmos de aprendizaje automático simbólico que pueden enlazar conjuntos de datos infiriendo consultas y mapeos adecuados. Nuestros objetivos principales se estructuran como sigue:

- Consulta de datos enlazados heterogéneos. Desarrollamos nuevos tipos de mapeos de esquemas para conjuntos de datos semi estructurados en formatos híbridos que incluyen bases de datos orientadas a grafos, colecciones RDF y bases de datos relacionales. Estos conducen a consultas recursivas sobre colecciones de datos enlazados para las que investigamos algoritmos de evaluación, problemas de análisis estático y aplicaciones concretas.

- Gestión de datos enlazados dinámicos. Para gestionar colecciones de datos enlazados dinámicos y flujos de trabajo, desarrollamos lenguajes de programación distribuidos centrados en los datos con flujos y paralelismo, basados en novedosos algoritmos para la respuesta a consultas incrementales. Asimismo, estudiamos la propagación de actualizaciones de datos dinámicos a través de mapeos de esquemas e investigamos métodos de análisis estático para flujos de trabajo de datos enlazados.

- Enlace de grafos. Por último, desarrollamos algoritmos de aprendizaje automático simbólico para inferir consultas y mapeos entre colecciones de datos enlazados en varios formatos de grafos a partir de ejemplos anotados.

El desarrollo de aplicaciones sobre estas tecnologías

Todos los equipos mencionados en esta sección desarrollan aplicaciones basadas en el conocimiento. El último equipo presentado, DYLISS, se dedica exclusivamente a la bioinformática. Las tecnologías, cada vez más potentes (por ejemplo, el análisis de secuencias), han acelerado el avance hacia un mapa completo del proceso biológico a nivel molecular y celular. El conocimiento representado en estos modelos biológicos debe ser compartido (entre herramientas de software y posteriormente entre el software y los usuarios) de manera que se preserve la semántica del conocimiento. La estandarización del conocimiento, en particular

sobre las regulaciones biológicas que son muy complejas de unificar (formato BioPAX), y la utilización de las numerosas bases de conocimiento disponibles (Reactome, Rhea, pathwaysCommnons, etc.) garantizarán una interoperabilidad semántica fiable.

DYLISS

Dinámica, lógica e inferencia para sistemas y secuencias biológicas

Las ciencias experimentales están pasando por una revolución de datos debido a la multiplicación de sensores que permiten medir la evolución de miles de componentes físicos o biológicos interdependientes a lo largo del tiempo. Cuando las mediciones son lo suficientemente precisas y variadas, pueden integrarse en un marco de aprendizaje automático para destacar las entidades mejor clasificadas dentro de los conjuntos de datos considerados. Sin embargo, **el interés biológico yace en la explicación de la clasificación, más precisamente en la identificación de los procesos biológicos que conducen a la especificidad de las entidades seleccionadas** con respecto al fenotipo considerado. Para ello es necesario tener en cuenta el conocimiento de dominio existente sobre las cadenas de compuestos biológicos implicados en las fuentes de datos, junto con sus reguladores.

Esto plantea varias cuestiones: en primer lugar, tenemos que **integrar las distintas fuentes de datos específicas del proyecto**, tanto juntas como con los datos del dominio de referencia y las bases de conocimiento. En segundo lugar, tenemos que **extraer modelos que apoyen la explicación** del papel de las entidades de interés, que tienen que ser coherentes con el conocimiento del dominio.

Lo más importante es que, aunque podamos adquirir cantidades de datos sin precedentes, éstos siguen sin estar a la altura de la complejidad biológica. Esto da lugar a un gran número de modelos (incluso considerando sólo los mínimos), todos ellos igualmente compatibles con las observaciones y el conocimiento del dominio. Evitar el sesgo de los enfoques codiciosos y la parcialidad observacional plantea una tercera cuestión: **considerar la familia exhaustiva de modelos consistentes** y ayudar a los expertos del dominio para explorarlos y analizarlos.

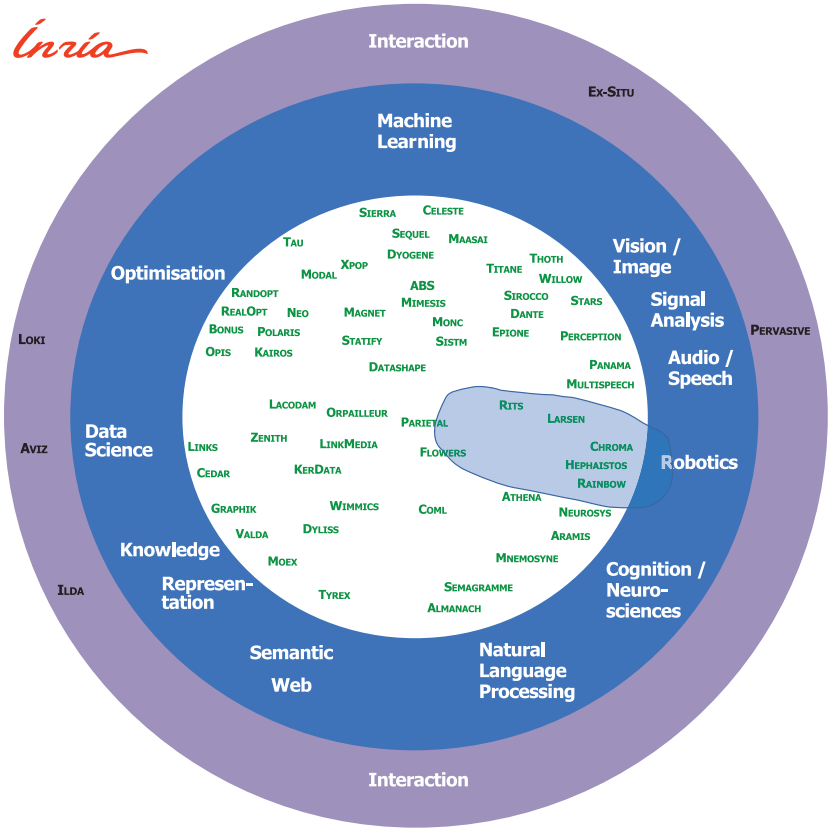
Para abordar estas cuestiones, DYLISS desarrolla métodos de análisis de datos y razonamiento basados en el conocimiento. Un primer eje es el desarrollo de métodos de estructuración e integración de datos para unificar fuentes de datos y corpus de conocimiento en grafos de conocimiento. Esto se apoya en las tecnologías de la Web Semántica y en los recursos de la iniciativa Linked Open Data (más de 1.600 repositorios de conocimiento para las ciencias de la vida). Un segundo eje consiste en aprovechar los datos estructurados para extraer familias de modelos que expliquen explícitamente el papel de las moléculas: esto se consigue con una combinación de métodos de aprendizaje a partir de ejemplos, enfoques basados en consultas y métodos de programación lógica que implican restricciones de sistemas dinámicos vistas como reglas de optimización. En el tercer eje, estos métodos también ayudan a los expertos del dominio a explorar y analizar exhaustivamente la familia de modelos.

5.6 Robótica y vehículos autónomos

La robótica combina muchas ciencias y tecnologías, desde el "nivel base" de la mecánica, la mecatrónica, la electrónica, el control, hasta el "nivel superior" de la percepción, la cognición, la colaboración y el razonamiento. En esta sección, aunque la inteligencia artificial en la robótica podría implicar ahondar en las funciones de nivel base para algunas características de procesamiento, sólo nos ocuparemos de los niveles superiores, los que se relacionan directamente con el campo de la IA.

Los recientes avances de la robótica son impresionantes. Los robots humanoides pueden caminar, correr, moverse en entornos conocidos y desconocidos y realizar tareas sencillas como agarrar objetos o manipular dispositivos. Los robots bio-inspirados son capaces de imitar los comportamientos de una gran cantidad de criaturas vivas muy diversas (insectos, pájaros, reptiles, roedores, etc.) y utilizar estos comportamientos para resolver eficazmente problemas complejos. El robot bípedo Atlas de Boston Dynamics (http://www.bostondynamics.com/robot_Atlas.html), mediante una percepción sencilla y mecanismos de control eficientes, puede desplazarse con eficacia en terrenos accidentados al aire libre y transportar objetos pesados, siguiendo al robot de cuatro patas BigDog de la misma empresa.

En relación con el aspecto cognitivo, gracias a los avances en el procesamiento del habla, la visión y la comprensión de escenas a partir de muchos sensores, y como consecuencia de las capacidades de razonamiento implementadas, los robots pueden tocar música, dar la bienvenida a los visitantes en los centros comerciales y conversar con los niños, entre otras cosas. Gracias a las funciones de coordinación entre una flota de robots, son capaces de jugar al fútbol juntos, pero ningún equipo de robots es aún apto para vencer a un equipo de humanos poco hábiles. Los vehículos autónomos son capaces de comportarse de forma segura durante largos períodos de tiempo y algunos países y estados de los EE.UU. podrían permitirles circular por las carreteras públicas en un futuro próximo, aunque todavía quedan muchas cuestiones pendientes, incluidas las éticas.



Los retos que abordan los equipos de Inria que desarrollan investigaciones sobre robots y vehículos de autoconducción son (i) la comprensión de la situación a partir de una entrada multisensorial; (ii) el razonamiento en condiciones de incertidumbre y la resiliencia; y (iii) la combinación de varios enfoques para la toma de decisiones. Para un análisis más profundo de los vehículos autónomos y conectados, consulte el libro blanco de Inria[28] (en francés), que indica que los coches totalmente autónomos no serán de uso generalizado antes del año 2040.

28. https://www.inria.fr/sites/default/files/2019-10/inrialivreblancvac-180529073843.pdf

Comprensión de situaciones a partir de información multisensorial

Para que un robot se mueva por zonas desconocidas, para que un coche autodirigido se pueda desplazar en el tráfico y para un robot de asistencia personal como Toi.Net (véase la sección 1), es esencial percibir el entorno y caracterizar la situación. Para ello, se utiliza la información procedente de múltiples sensores (como por ejemplo, visión, láser, sonido, Internet, o datos road2car en el caso de los vehículos). Las situaciones pueden ser simples símbolos, ontologías o representaciones más sofisticadas de actores y objetos presentes en un entorno. Una buena caracterización de la situación puede ayudar al robot a tomar decisiones, incluso en algunos casos a infringir la ley o un reglamento para salvar la vida de los pasajeros del vehículo.

Razonamiento bajo incertidumbre, resiliencia

Los robots actúan en el mundo físico y tienen que enfrentarse a fallos y defectos de muchos tipos: cortes de red, sensores defectuosos, peligros electrónicos, etc. Algunos sensores proporcionan información incompleta o tienen márgenes de error que generan incertidumbre en los datos. Sin embargo, un robot móvil autónomo debe realizar su operación de forma continua, sin intervención humana y durante largos periodos de tiempo. Un desafío para las arquitecturas y el software de los robots es lidiar con información incierta o ausente y con aquella sólo disponible en distintos momentos de adquisición. Los algoritmos anytime (en cualquier momento) que proporcionan una respuesta bajo petición pueden ser una solución en el caso de que se necesite realizar una toma de decisiones rápida, aunque ésta no sea perfecta.

Combinar varios enfoques para la toma de decisiones

Un robot puede disponer de una gran variedad de datos e información para tomar una decisión. Los datos procedentes de diferentes sensores, información sobre el entorno en la forma de una evaluación de la situación, recuerdos de decisiones tomadas en el pasado, reglas y normas implementadas en la memoria del robot, son hechos y datos respecto a los cuales existe una necesidad de combinarlos y realizar un razonamiento híbrido a partir de datos numéricos, continuos o discretos, como también de representaciones semánticas. Además, como se ha visto anteriormente, este razonamiento también debe tener en cuenta la incertidumbre, por lo que la investigación sobre la toma de decisiones para robots tiene que abordar este reto. Una posible solución es el aprendizaje automático no supervisado y el aprendizaje por refuerzo de situaciones e interpretaciones semánticas.

Colaboración entre humanos y robots

En la mayoría de las situaciones de la vida real –como la asistencia a personas mayores, la conducción autónoma o el funcionamiento en fábricas– los robots deben interactuar adecuadamente con los usuarios y operadores humanos. Esta interacción es necesaria en ambos sentidos: obviamente, para que los robots entiendan los objetivos y acciones de los humanos (véase, por ejemplo, el libro de Stuart Russell sobre el tema)[29], pero también para que los humanos entiendan los objetivos y acciones que realizan los robots en su presencia. Un buen ejemplo de esto último se encuentra en un informe sobre la seguridad de la conducción automatizada publicado por un consorcio de partes interesadas, entre las que se encuentran los principales fabricantes alemanes[30], en el que se afirma que: "La HMI [interfaz hombre-máquina] debe diseñarse cuidadosamente para tener en cuenta los rasgos y estados psicológicos y cognitivos de los seres humanos, con el objetivo de optimizar la comprensión de la tarea y la situación por parte del ser humano y de reducir el mal uso accidental o las operaciones incorrectas".

HEPHAISTOS

Hexápodo, fisiología, asistencia y robótica

El objetivo del proyecto HEPHAISTOS es establecer una metodología genérica para el diseño y la evaluación de un ecosistema de asistencia adaptable e interactivo para las personas mayores y vulnerables, que proporcione además asistencia a los ayudantes, solicitudes de datos médicos y pueda gestionar situaciones de emergencia. Más concretamente, nuestros objetivos son desarrollar dispositivos con las siguientes propiedades:

- Pueden adaptarse al usuario final y a su entorno cotidiano;
- Deben ser asequibles y mínimamente intrusivos;
- Pueden controlarse a través de una gran variedad de interfaces sencillas; y
- Pueden llegar a utilizarse para controlar el estado de salud del

29. Stuart Russell. Human compatible, AI and the problem of control. Penguin Books, 2019.
30. Safety first for Automated Driving, Aptiv, BMW, Baidu Continental, Daimler et al., julio de 2019.

usuario final con el fin de detectar patologías emergentes.

La asistencia se proporcionará a través de una red de dispositivos de comunicación que pueden estar diseñados específicamente para esta tarea o ser simplemente una adaptación/instrumentación de objetos de la vida cotidiana.

La población a la que se dirige se limita a aquellas personas con problemas de movilidad (en aras de la simplicidad, esta población se denominará "personas mayores" en lo sucesivo en este libro blanco, si bien nuestro trabajo también se ocupa de una variedad de personas -por ejemplo, personas discapacitadas o lesionadas-) y los dispositivos de asistencia tendrán que apoyar la autonomía individual (en casa y al aire libre), proporcionando recursos complementarios en relación con las capacidades existentes de la persona. La personalización y adaptabilidad son factores clave de éxito y aceptación. Nuestro objetivo a largo plazo será proporcionar dispositivos robotizados de asistencia, incluidos objetos inteligentes, que puedan ayudar a las personas discapacitadas, mayores y minusválidas en su vida personal.

La asistencia en este sentido es un campo muy amplio y un solo equipo-proyecto no puede abordar todos los temas y problemas relacionados con la misma. Por ello, HEPHAISTOS se centrará en los siguientes retos sociales principales:

- **La movilidad:** las entrevistas y observaciones previas obtenidas en el equipo de HEPHAISTOS han demostrado que ésta era una de las principales preocupaciones de todos los actores del ecosistema. La movilidad es un factor clave para mejorar la autonomía personal y reforzar la privacidad, la percepción de autonomía y autoestima.

- **Gestión de situaciones de emergencia:** las situaciones de emergencia (por ejemplo, una caída) pueden tener consecuencias dramáticas para las personas mayores. Lo ideal sería que los dispositivos de asistencia fueran capaces de prevenir estas situaciones y, al menos, de detectarlas con el fin de enviar una alarma y minimizar los efectos sobre la salud de las personas mayores.

- **Seguimiento médico:** las personas mayores pueden tener una

trayectoria de vida que cambia rápidamente y la comunidad médica carece de información sintética oportuna sobre esta evolución, mientras que las tecnologías disponibles permiten obtener información bruta de forma no intrusiva y de bajo costo. Pretendemos proporcionar indicadores sintéticos de salud que tengan en cuenta las incertidumbres de las mediciones, obtenidos a través de una red de dispositivos de asistencia. Sin embargo, el respeto a la privacidad de la vida, la protección de las personas mayores y las consideraciones éticas imponen garantizar la confidencialidad de los datos y un estricto control de dicho servicio por parte de la comunidad médica.

- **Rehabilitación y biomecánica:** nuestros objetivos en materia de rehabilitación son: 1) proporcionar indicadores más objetivos y robustos, que tengan en cuenta las incertidumbres de las mediciones para evaluar el progreso de un proceso de rehabilitación; 2) proporcionar procesos y dispositivos (incluido el uso de la realidad virtual) que faciliten un proceso de rehabilitación y sean más flexibles y fáciles de usar tanto para los usuarios como para los médicos. La biomecánica es una herramienta esencial para evaluar la pertinencia de estos indicadores, para acceder a parámetros fisiológicos difíciles de medir directamente y para preparar eficazmente experimentos en la vida real.

MARIONET-ASSIST, robot paralelo accionado por cable para asistir a personas con movilidad reducida - © Inria _ Photo H. Raguet

LARSEN

Autonomía de por vida y habilidades de interacción para robots en un entorno de detección

El equipo de Larsen pretende combinar los recientes avances en inteligencia artificial, aprendizaje automático y toma de decisiones con los de la robótica para diseñar robots más inteligentes, flexibles y capaces de cooperar con los humanos. El objetivo es ir más allá de la robótica tradicional, que se limita a tareas repetitivas en entornos muy controlados en los que los humanos tienen poca cabida.

Para lograr este objetivo, el equipo está desarrollando métodos para dotar a los robots de habilidades de autonomía a largo plazo, que les permitan operar las 24 horas del día, y de habilidades que les permitan interactuar de forma natural con los humanos teniendo en cuenta los sensores integrados y externos del entorno.

El equipo se beneficia de una rica infraestructura de pruebas: un departamento equipado con sensores, un escenario robótico con captura de movimiento, otro de vuelo para drones con captura de movimiento y muchos robots, incluyendo robots humanoides iCub y Talos, un cuadrúpedo, dos hexápodos, dos manipuladores móviles y dos manipuladores industriales, etc.

El objetivo de Larsen es diseñar robots que tengan la capacidad de:

- Manejar un entorno dinámico y situaciones imprevistas;
- Hacer frente a los daños físicos;
- Interactuar física y socialmente con los humanos;
- Colaborar entre sí;
- Explotar la multitud de mediciones de los sensores de su entorno;
- Mejorar su aceptabilidad y usabilidad por parte de los usuarios finales que no cuentan con conocimientos de robótica.

Todas estas capacidades pueden resumirse en los dos objetivos siguientes:

- Autonomía de por vida: realizar tareas de forma continua adaptándose a los cambios repentinos o graduales tanto del entorno como de la morfología del robot;

- Interacción natural con los sistemas robóticos: interactuar con otros robots y con los humanos durante largos periodos de tiempo, teniendo en cuenta que las personas y los robots aprenden unos de otros cuando conviven juntos.

Brazo robótico de Creativ'Lab - © Inria _ Photo D. Betzinger

RAINBOW

Robótica basada en sensores e interacción humana

La visión a largo plazo del equipo Rainbow es desarrollar la próxima generación de robots basados en sensores capaces de navegar y/o interactuar en entornos complejos no estructurados junto a usuarios humanos. Evidentemente, la palabra "juntos" puede tener significados muy diferentes según el contexto: por ejemplo, puede referirse a la mera

coexistencia (los robots y los humanos comparten un espacio mientras realizan tareas independientes), a la conciencia humana (los robots necesitan conocer el estado y las intenciones de los humanos para ajustar adecuadamente sus acciones) o a la cooperación real (los robots y los humanos realizan alguna tarea compartida y necesitan coordinar sus acciones).

Tal vez se podría debatir que estos dos objetivos entran en conflicto, ya que una mayor autonomía del robot debería implicar una menor (o nula) intervención humana. Sin embargo, creemos que nuestro enfoque general de la investigación está bien motivado, ya que:

- A pesar de los numerosos avances en la autonomía de los robots, las decisiones complejas y de alto nivel basadas en la cognición siguen estando fuera de su alcance. En la mayoría de las aplicaciones que implican tareas en entornos no estructurados, incertidumbre e interacción con el mundo físico, la asistencia humana sigue siendo necesaria y probablemente lo será durante las próximas décadas. Por otro lado, los robots son extremadamente capaces de ejecutar de forma autónoma tareas específicas y repetitivas, con gran velocidad y precisión, y de operar en entornos peligrosos/remotos, mientras que los humanos poseen unas capacidades cognitivas y una conciencia del mundo inigualables que les permiten tomar decisiones complejas y rápidas;

- La cooperación entre humanos y robots es a menudo una restricción implícita de la propia tarea robótica. Pensemos, por ejemplo, en el caso de los robots de asistencia que ayudan a pacientes lesionados durante su recuperación física, o en los dispositivos de aumento humano. Por eso es importante estudiar formas adecuadas de implementar esta cooperación;

- Por último, las normas de seguridad pueden exigir la presencia en todo momento de una persona encargada de supervisar y, en caso necesario, tomar el control directo de los trabajadores robóticos. Por ejemplo, este es un requisito común en todas las aplicaciones que implican tareas en espacios públicos, como los vehículos autónomos en espacios concurridos, o incluso los vehículos aéreos no tripulados cuando vuelan en el espacio aéreo civil, como sobre zonas urbanas o pobladas.

Dentro de este panorama general, las actividades de Rainbow se centrarán especialmente en el caso de la **cooperación (compartida) entre robots y humanos**, en la búsqueda de la siguiente visión: por un lado, dotar a los robots de un amplio grado de autonomía para que puedan operar eficazmente en entornos no triviales (por ejemplo, fuera de entornos de fábricas completamente definidos). Por otro lado, incluir a los usuarios humanos en el círculo para que tengan el control (parcial y bilateral) de algunos aspectos del comportamiento general del robot. Tenemos previsto abordar estos retos desde las perspectivas **metodológica, algorítmica y de aplicación.** Los principales ejes de investigación a lo largo de los cuales se articulan las actividades de Rainbow son tres pilares de apoyo **(detección óptima y consciente de la incertidumbre; control avanzado basado en sensores; y háptica para aplicaciones robóticas)** que pretenden desarrollar métodos, algoritmos y tecnologías para hacer realidad el tema central del **control compartido de sistemas robóticos complejos.**

Moviendo una silla de ruedas inteligente en la realidad virtual - © Inria _ Photo G. Scagnelli

Vehículos autónomos

Los primeros problemas fundamentales en relación con el uso de la IA en el ámbito de los vehículos autónomos (AV por sus siglas en inglés) son los de la explicabilidad y la coherencia de los resultados del algoritmo. Estos son el requisito preliminar para el desarrollo de los marcos legales necesarios para las pruebas a gran escala y el despliegue de los AV en las redes de carreteras y ciudades reales. En el plano técnico, los primeros retos son los costos computacionales y el consumo de energía si se despliegan ampliamente las arquitecturas de IA dedicadas (tarjetas y otras).

Otros retos algorítmicos están relacionados con la necesidad de contar con un gran número de multi-sensores y conjuntos de datos de múltiples escenarios anotados. En los últimos años, el esfuerzo global por publicar una investigación reproducible ha dado lugar a un número creciente de códigos abiertos y conjuntos de datos públicos, allanando el camino para obtener resultados interesantes. En el año 2012, KITTI fue el primer conjunto de datos a gran escala para la conducción autónoma con visión y, a partir de entonces, conjuntos de datos públicos como ScanNet (2018) para el procesamiento en 3D, nuScenes (2019) para la conducción multisensorial, SemanticKITTI (2019) para las escenas de conducción en 3D y muchos otros que permitieron colectivamente un gran salto en cuanto al rendimiento. Es más, muchas investigaciones mostraron el beneficio de pre-entrenar redes profundas en estos grandes conjuntos de datos públicos para una gran variedad de tareas, demostrando que las características de alto nivel pueden ser compartidas incluso para tareas de distintas naturalezas.

Aun así, la línea de investigación actual adolece al seguir este paradigma supervisado dado que requiere grandes conjuntos de datos (del orden de miles/millones de datos) cuya anotación es tediosa y rutinaria. Aunque el aprendizaje supervisado aporta sin duda el mejor rendimiento, el costo del etiquetado acabará siendo insostenible, ya que tanto el tamaño del conjunto de datos como el número de sensores aumentan constantemente. Sin olvidar que abarcar todas las condiciones (iluminación, escenarios de tráfico, climas, etc.) en un solo conjunto de datos es poco práctico. Por ejemplo, no existe ningún conjunto de datos disponible que abarque escenarios de conducción peligrosos. Es necesario aprovechar el aprendizaje semi o no supervisado para garantizar la escalabilidad y aplicabilidad de los algoritmos al mundo exterior real donde, en última instancia, se enfrentan a situaciones no vistas en el conjunto de entrenamiento. El "Santo Grial" de la inteligencia artificial general está, por ahora, fuera del alcance de nuestros conocimientos actuales, pero las técnicas prometedoras en el aprendizaje

por transferencia permiten ampliar la formación realizada de forma supervisada a nuevos conjuntos de datos sin etiquetar, por ejemplo con la adaptación de dominios. Experimentos emocionantes realizados por el equipo de RITS y otros laboratorios de investigación lograron demostrar la capacidad de aplicar dicha estrategia, por ejemplo, al aprendizaje por transferencia, a condiciones de luz cambiantes (el entrenamiento sobre datos diurnos y la realización de pruebas sobre datos nocturnos), al clima (conducción en condiciones despejadas o lluviosas), o incluso la naturaleza de los datos (simulador de conducción real).

Hoy en día, el aprendizaje automático se utiliza ampliamente en el campo de la virtualidad aumentada (AV) para los sistemas de percepción. Sin embargo, otras técnicas de IA parecen tan prometedoras como el aprendizaje automático, además de ser más fáciles de interpretar. La IA abre ciertamente el camino a nuevas áreas de investigación y demuestra una gran capacidad de resolver problemas de larga data cruciales para la conducción autónoma (por ejemplo, el etiquetado semántico de entornos exteriores complejos).

RITS

Robótica y sistemas de transporte inteligentes

El equipo-proyecto RITS forma parte de un proyecto multidisciplinario de Inria que trabaja en el campo de la robótica para sistemas de transporte inteligentes. En concreto, trata de combinar la inteligencia artificial y la modelización matemática para diseñar sistemas avanzados de robótica inteligente para la movilidad autónoma y sostenible.

Entre los temas científicos abordados se cuentan:

- Técnicas multimodales para la comprensión de la escena a partir de una cámara, datos láser, GPS, etc.;
- Entrenamiento no supervisado o débilmente supervisado (adaptación del dominio, destilación de datos);
- Control de bajo y alto nivel de vehículos;
- Toma de decisiones para la conducción autónoma;

- Modelización y simulación del tráfico a gran escala;
- Control y optimización de los sistemas de transporte por carretera;
- Desarrollo y despliegue de vehículos automatizados (coches cibernéticos, vehículos privados, etc.).

El objetivo de estos estudios es **mejorar el transporte por carretera** en cuanto a la seguridad, eficiencia y comodidad y también minimizar las molestias. El enfoque técnico se basa en la asistencia al conductor, llegando hasta la automatización total de la conducción. El equipo-proyecto pone a disposición de los distintos equipos asociados algunos medios importantes, como una flota de una decena de vehículos conducidos por computador, diversos sensores e instalaciones informáticas avanzadas, incluida una herramienta de simulación. Un sistema experimental basado en vehículos totalmente automatizados se ha montado en las instalaciones de Inria en Rocquencourt con fines demostrativos.

Una de las plataformas de conducción autónoma de RITS - B4 © Inria / Photo H. Raguet.

CHROMA

Navegación robótica cooperativa y con conciencia humana en entornos dinámicos

El objetivo general de Chroma es abordar cuestiones fundamentales y abiertas que se encuentran en la intersección de los campos de investigación emergentes denominados "Human Centred Robotics" (robótica centrada en el ser humano) [1]. Más concretamente, el objetivo es diseñar algoritmos y desarrollar modelos que permitan a los robots móviles navegar y cooperar en entornos dinámicos y poblados por humanos. Chroma participa en todos los aspectos de decisión relativos a las tareas de navegación de uno o varios robots, incluyendo la percepción y la planificación del movimiento.

El objetivo general es construir comportamientos robóticos que permitan a uno o varios robots operar con seguridad entre humanos en entornos parcialmente conocidos, donde el tiempo, la dinámica y las interacciones desempeñan un papel importante. Los recientes avances en la potencia de computación integrada, las tecnologías de sensores y comunicaciones y los sistemas mecatrónicos miniaturizados, hacen posible los avances tecnológicos necesarios (incluso desde el punto de vista de la escalabilidad).

Chroma se posiciona claramente en el tema de investigación "Artificial Intelligence and Autonomous Systems" (Inteligencia artificial y sistemas autónomos) del Plan Estratégico Inria 2018-2022. Más concretamente nos referimos al reto "Inteligencia aumentada" (vehículos autónomos conectados) y al reto "Mundo digital centrado en el ser humano" (adaptación interactiva).

Mini-UAV Crazyflies 2.0, controlado por banda ultra ancha UWB)
© Inria _ Photo C. Morel

[1] Montreuil, V.; Clodic, A.; Ransan, M.; Alami, R., "Planning Human Centred Robot Activities", en Systems, Man and Cybernetics, 2007

5.7 Neurociencias y cognición

La IA y el estudio de la cognición tienen una larga historia de colaboración. Los paradigmas de la IA suelen basarse en conceptos tomados de la investigación sobre la cognición y, a su vez, pueden contribuir a los avances en la ciencia de la cognición. Por ejemplo, la realización de experimentos con grandes redes neuronales puede servir como una herramienta para que los neurocientíficos comprueben nuevos modelos del cerebro. La intersección entre la IA, las neurociencias y la cognición fue el incentivo de algunos de los mayores proyectos de investigación emprendidos por la humanidad, como el Human Brain Project Flagship, financiado por la Comisión Europea, o la BRAIN Initiative del NIH en Estados Unidos.

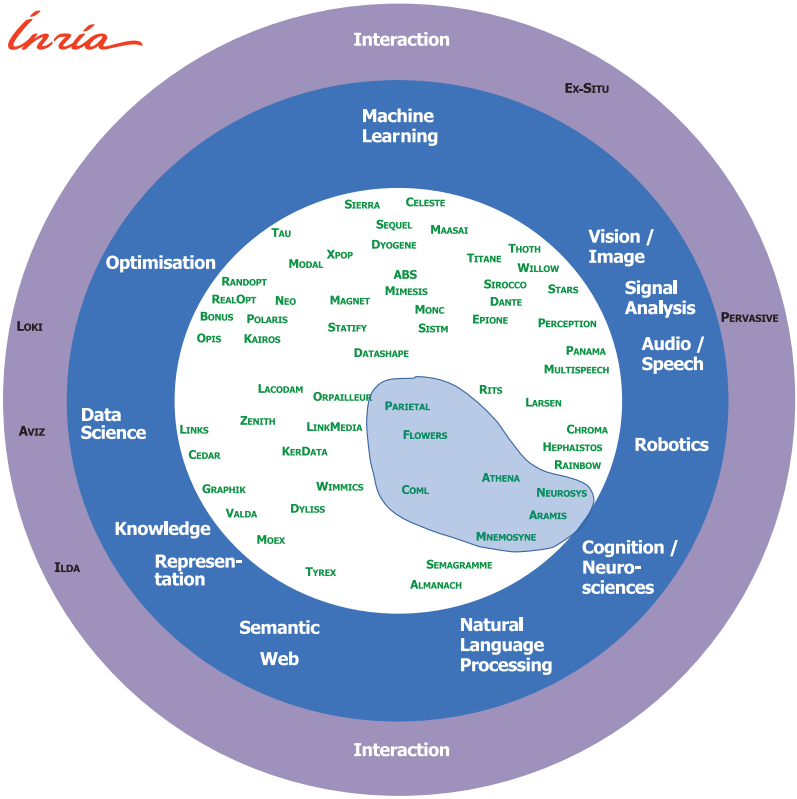
Una tendencia emergente en la IA es seguir la propuesta del premio Nobel Daniel Kahneman de modelar el pensamiento humano como la interacción continua de dos sistemas, el Sistema 1 y el Sistema 2.

Del libro de Kahneman, Pensar rápido, pensar despacio:

El pensamiento del Sistema 1 es RÁPIDO, AUTOMÁTICO, ocurre INCONSCIENTEMENTE y requiere un ESFUERZO MÍNIMO.

El pensamiento del Sistema 2 es más LENTO, requiere ESFUERZO y se produce de forma CONSCIENTE y DELIBERADA.

La mayoría de los sistemas de aprendizaje automático que utilizan redes neuronales pueden asignarse al Sistema 1 como, por ejemplo, el caso de la visión, el reconocimiento del habla, la conducción autónoma, etc. La pregunta de cómo desarrollar las capacidades del Sistema 2 es objeto de debate: algunos autores creen que estas capacidades pueden obtenerse utilizando modelos más sofisticados del cerebro, es decir, redes neuronales más complejas; otros están convencidos de que enfoques complementarios de la IA, como el razonamiento semántico y en base al conocimiento, serán útiles para este fin. A mediados del año 2020, este debate seguía en pañales; se requiere una investigación y experimentación más extensa y esto llevará años, si no décadas.



Dentro de Inria, algunos equipos de investigación se encuentran en la intersección de la IA y las neurociencias. Su trabajo puede definirse como contribuciones al pensamiento del Sistema 1 y del Sistema 2, aunque algunos puedan estar más relacionados con uno de ellos.

Sus principales retos científicos son los siguientes:

Construir mejores modelos del cerebro

Este reto es compartido por todos los equipos de este ámbito, ya que es el problema más fundamental de las neurociencias y la cognición. Puede afectar tanto al cerebro sano como a las enfermedades cerebrales. Para ello, se explotan varios paradigmas de modelización y se combinan con diversos datos, como los obtenidos del MRI (imagen por resonancia magnética), el EEG (electroencefalograma) y el

MEG (magneto-encefalografía). Se desarrollan modelos para células individuales, grupos de células, estructuras de conectividad y patrones de actividad almacenados en bibliotecas.

Un camino hacia el sentido común

El razonamiento de sentido común es una motivación primordial para la IA. Sigue siendo un objetivo lejano para todos los enfoques, incluso después de grandes inversiones y años de investigación, como el proyecto CYC[31] de Doug Lenat en la década de 1990. La investigación en neurociencias y cognición puede, en última instancia, aportar nuevos conocimientos sobre el razonamiento humano de sentido común, pero nuestra historia no tan reciente invita a cierta modestia al respecto.

Acceso a las funciones/autonomía ejecutivas de orden superior

Las funciones ejecutivas superiores (organización temporal del comportamiento, capacidad de generalización, manipulación del conocimiento implícito y explícito, etc.), así como la autonomía real (aprendizaje continuo, flexibilidad, aprendizaje con uno o pocos ejemplos) siguen siendo retos importantes que sólo estamos empezando a abordar.

ARAMIS

Algoritmos, modelos y métodos para imágenes y señales del cerebro humano

Actualmente es posible medir múltiples características de las enfermedades cerebrales en pacientes vivos gracias a los enormes avances de las tecnologías de neuroimagen, genómica y biomarcadores. La recopilación de datos multimodales en grandes bases de datos de pacientes ofrece una visión completa de las alteraciones cerebrales, los procesos biológicos, factores de riesgo genético y síntomas. Un reto importante es ahora **construir modelos numéricos de enfermedades cerebrales a partir de datos multimodales de pacientes**, en función del desarrollo de enfoques específicos basados en datos. Dichos modelos ayudarán a profundizar nuestro conocimiento de las enfermedades neurológicas y a

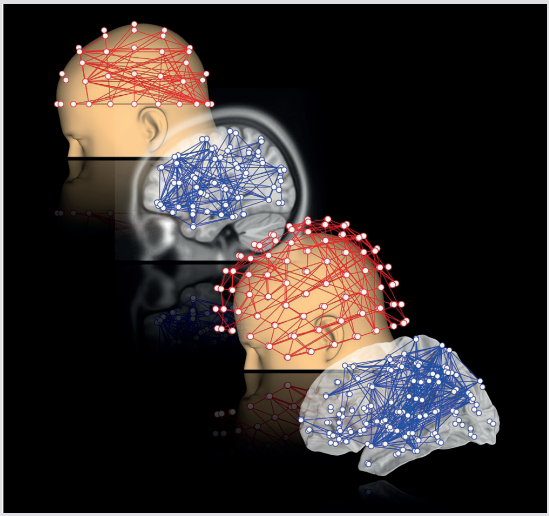
31. https://en.wikipedia.org/wiki/Cyc

diseñar sistemas eficaces de asistencia a la toma de decisiones clínicas.

El objetivo del equipo-proyecto ARAMIS de Inria es diseñar nuevos planteamientos de aprendizaje automático y análisis de datos para modelar enfermedades cerebrales y sistemas de apoyo a la toma de decisiones para ayudar a los médicos. Para ello, desarrollamos enfoques que puedan integrar múltiples tipos de datos adquiridos del paciente vivo, incluyendo neuroimágenes, biomarcadores periféricos, datos clínicos y ómicos. Una primera línea de investigación está dedicada a la detección de alteraciones en los datos de imágenes cerebrales y al diseño de sistemas de IA para asistir a los radiólogos [2]. Una segunda línea se refiere al análisis de fenómenos temporales a partir de datos longitudinales. Esto implica el desarrollo de sofisticados modelos de efectos mixtos utilizando herramientas de la geometría de Riemann [3]. Estos modelos pueden reconstruir escenarios de progresión de la enfermedad a nivel individual y poblacional. Se aplican en las herramientas informáticas Leaspy1 y Deformetrica2 disponibles gratuitamente. Un tercer eje pretende modelar las interacciones funcionales entre áreas cerebrales distantes que subyacen a los procesos cognitivos. Se basa en enfoques que pueden modelar la organización de redes cerebrales complejas [1]. Se aplican al diseño de nuevos dispositivos, interfaces cerebro-computadora y neurofeedback, para la rehabilitación de pacientes neurológicos. El equipo dedica muchos esfuerzos a la transferencia de estas herramientas a los estudios clínicos, mediante el desarrollo de la plataforma de software de Clinica.3 Por último, también proporcionamos directrices y marcos para la investigación reproducible en este campo. Tres integrantes del equipo (N. Burgos, O. Colliot, S. Durrleman) son catedráticos del PRAIRIE 3IA Institute.

[1] De Vico Fallani F, Richiardi J, Chavez M, y Achard S, Graph analysis of functional brain networks: practical issues in translational neuroscience., Philosophical Transactions of the Royal Society of London Series B, Biological Sciences, 369:1653, 2014.
[2] Samper-González J, Burgos N, Bottani S, Fontanella S, Lu P, Marcoux A, Routier A, Guillon J, Bacci M, Wen J, Bertrand A, Bertin H, Habert M-O, Durrleman S, Evgeniou T, y Colliot O, Reproducible evaluation of classification methods in Alzheimer's disease: Framework and application to MRI and PET data., NeuroImage, 183, 504–521, 2018
[3] Schiratti J-B, Allassonnière S, Colliot O, y Durrleman S, A Bayesian Mixed-Effects Model to Learn Trajectories of Changes from Repeated Manifold-Valued Observations., Journal of Machine Learning Research, 18:133, 1–33, 2017

1 https://gitlab.com/icm-institute/aramislab/leaspy
2 https://www.deformetrica.org/
3 http://www.clinica.run



Análisis de la red de conexiones complejas en el cerebro
© Inria _ Fabrizio De Vico Fallani - ARAMIS

ATHENA

Imágenes computacionales del sistema nervioso central

Aunque en las últimas décadas se han conseguido avances excepcionales en la exploración del cerebro humano, éste sigue siendo terra-incognita y exige esfuerzos de investigación específicos para comprender mejor su arquitectura y funcionamiento.

El equipo-proyecto ATHENA tiene como objetivo general comprender mejor la estructura y la función del cerebro humano mediante el desarrollo de una nueva generación de modelos computacionales y avances metodológicos para la cartografía de la conectividad cerebral. Para resolver la visión limitada del cerebro que proporciona una sola modalidad de imagen, y recuperar la conectividad estructural y funcional del cerebro, los modelos construidos por el equipo se basan firmemente en modalidades avanzadas y complementarias de imagen integrada no invasiva e in vivo: la Resonancia Magnética de Difusión (Magnetic

Resonance Imaging o dMRI) y la electro y magneto-encefalografía (EEG y MEG).

Las principales líneas de investigación del equipo son :

1. Desarrollar herramientas matemáticas y computacionales rigurosas para la adquisición, procesamiento y análisis combinado de datos de dMRI y MEG y EEG.

2. Impulsar el estado de la técnica en el mapeo computacional de la conectividad cerebral y las interfaces cerebro-computadora (Brain Computer Interfaces o BCI).

3. Desarrollar y abordar, con nuestros colaboradores, aplicaciones clínicas y de BCI.

Esto ayudará en gran medida a comprender y reconstruir mejor la conectividad estructural y funcional del cerebro y a proporcionar un valor clínico añadido para identificar y caracterizar mejor las anomalías en la conectividad cerebral. Aunque la BCI se defiende como medio para comunicar y ayudar a recuperar la movilidad o la autonomía en casos muy graves de pacientes discapacitados, también es una nueva herramienta para sondear y entrenar de forma interactiva el cerebro humano.

Un tercio de la carga de todas las enfermedades en Europa se debe a problemas causados por patologías que afectan al cerebro. Los objetivos de ATHENA representan un fantástico reto científico, así como una necesidad clínica acuciante que, cuando se resuelva, tendrá un impacto positivo en la inaceptable carga de las enfermedades cerebrales y abrirá nuevas perspectivas en la neurociencia.

Mapeo del cerebro - © Inria _ Photo C. Morel

MNEMOSYNE

Sinergia mnemónica

En la frontera entre la Inteligencia Artificial y la neurociencia computacional, el equipo de MNEMOSYNE se propone modelar las principales formas de memoria y aprendizaje en el cerebro y estudiar cómo se organizan e implementan funciones cognitivas complejas. En la neurociencia, se señala una importante dicotomía entre la memoria y el aprendizaje explícitos (por ejemplo, semánticos, episódicos) e implícitos (por ejemplo, procedimentales, habituales). Los mecanismos clave para entender funciones cognitivas como el razonamiento, la toma de decisiones, los procesos atencionales y el lenguaje se basan en la competencia, cooperación y transferencia entre estas distintas formas de aprender y memorizar información, las que actualmente son objeto de grandes avances en diferentes campos de la neurociencia.

El equipo de MNEMOSYNE diseña modelos de las estructuras y circuitos neuronales subyacentes bajo esta visión funcional de la organización y la dinámica del cerebro. Los modelos se basan en diferentes tipos de arquitecturas neuronales (a futuro, recurrentes, convolucionales y generativas) con el reto de imitar los bucles entre el córtex prefrontal y los ganglios basales, y sus interacciones con el córtex sensorial, el hipocampo, la amígdala y otras estructuras cerebrales, identificadas como el sustrato de las funciones cognitivas objetivo. Estos modelos sirven de base para las colaboraciones del equipo con las comunidades neurocientíficas y médicas; también son la base de su posicionamiento original en el Aprendizaje Automático, hacia la Inteligencia Artificial General. El equipo considera un reto importante proponer modelos computacionales, plasmados en agentes virtuales o reales que interactúen en línea con el entorno y sean capaces de extraer de forma autónoma estructuras para construir un modelo distribuido del mundo, seleccionar con flexibilidad la mejor estrategia para alcanzar objetivos internos y externos y aprender de sus errores.

Los temas de investigación recientes corresponden a la adquisición del lenguaje y la extracción de la sintaxis, la codificación de objetivos en el comportamiento motivado, la transferencia del comportamiento

dirigido a objetivos al habitual, la planificación y el razonamiento con una memoria funcional y la deliberación retrospectiva y prospectiva. Estos modelos se construyen en estrecha interacción con neurocientíficos, en asociación con protocolos experimentales, y se explotan para considerar casos patológicos en el ámbito médico. Asimismo, se trasladan al mundo socioeconómico con aplicaciones industriales y se investiga activamente su impacto en las ciencias sociales y las humanidades, especialmente en proyectos conjuntos con las ciencias de la educación, la lingüística, economía y filosofía.

PARIETAL

Modelado de la estructura, función y variabilidad del cerebro a partir de datos de MRI de alto campo

La inteligencia artificial es un campo polifacético y el estudio del cerebro mediante imágenes cerebrales ofrece una oportunidad casi única de explorar estas diferentes facetas. El equipo Parietal, integrante de la mayor plataforma francesa de imágenes cerebrales, Neurospin, explora los vínculos entre el cerebro, imágenes y la cognición.

En primer lugar, los datos adquiridos sobre el cerebro se proporcionan en forma de señales (registros electrofisiológicos) o de imágenes, como aquellas adquiridas en la Imagen por Resonancia Magnética. La explotación correcta de estos datos implica problemas estadísticos y de estimación a gran escala, que hoy en día se resuelven mediante métodos de optimización y aprendizaje estadístico (aprendizaje automático o machine learning), uno de los ámbitos de la IA. Por ejemplo, la reconstrucción de la actividad eléctrica cerebral a partir de las mediciones de los campos electromagnéticos tomados en la superficie del cuero cabelludo, requiere la solución de un **problema inverso mal planteado**, para el que las herramientas de regresión a gran escala ofrecen soluciones óptimas. El equipo de Parietal ha desarrollado modelos y algoritmos especialmente eficaces para una regresión parsimoniosa. Del mismo modo, reconstruir una imagen de resonancia magnética del cerebro a partir de un número limitado de mediciones para reducir el tiempo de adquisición, equivale a resolver un problema inverso formalmente

similar. Para estos dos problemas, los investigadores de Parietal desarrollan métodos basados en el aprendizaje profundo, que conducen a solucionadores más rápidos para el análisis a gran escala.

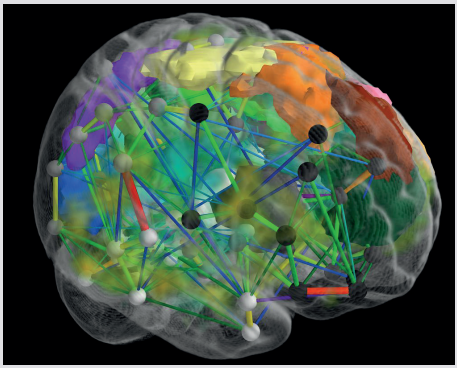
Por otro lado, a veces es necesario extraer patrones presentes en los datos de la actividad cerebral para construir modelos mucho más sencillos de la información basada en estos patrones. El equipo de Parietal ha desarrollado **técnicas de aprendizaje de diccionarios** y, trabajando sobre la estructura de los estimadores, ha desarrollado algoritmos muy eficientes que pueden analizar millones de imágenes del cerebro en un tiempo razonable. El mismo método permite también extraer patrones de series temporales.

En otros aspectos metodológicos, se está trabajando en el análisis de las **garantías estadísticas**: cuando se sostiene que la actividad de una región del cerebro predice el comportamiento de una persona, ¿cómo garantizar que es así y que no se trata de una interpretación errónea? Es difícil demostrar que una región determinada desempeña un papel en la predicción cuando muchas otras áreas podrían tener el mismo efecto. Los investigadores de Parietal desarrollan técnicas para hallar intervalos de confianza que permitan establecer que las relaciones estadísticas destacadas en las imágenes son realmente fiables.

Las imágenes funcionales del cerebro representan la activación cuando el sujeto realiza determinadas tareas, como por ejemplo ver una película. Sin embargo, aunque es complicado describir con detalle las operaciones mentales que se suceden al ver una película o escuchar una historia, ahora disponemos de redes neuronales artificiales que lo hacen tan bien o incluso mejor que los humanos. Por eso es apasionante estudiar si **ciertas regiones del cerebro podrían reaccionar como neuronas artificiales**. Los investigadores de Parietal han demostrado que ciertas áreas de la corteza visual se comportan como capas sucesivas de una red neuronal profunda. Ahora estamos estudiando si los sistemas modernos de procesamiento del lenguaje pueden explicar la respuesta observada en el cerebro al escuchar una historia.

El conocimiento del cerebro no se detiene en el procesamiento de imágenes y señales: los experimentos producen resultados que deben integrarse en **bases de conocimiento**, de modo que puedan

incorporarse a teorías unificadoras o puedan reutilizarse para analizar mejor nuevos datos. Hasta ahora, este trabajo se ha realizado mediante la lectura de publicaciones en la materia. La reciente investigación de Parietal ha contribuido a automatizar la adquisición y el uso del conocimiento de las publicaciones (neuroquery.org), pero también a probar los resultados de varias decenas de experimentos de neurociencia cognitiva para integrarlos en un modelo. De este modo, podemos sintetizar la información experimental recogida en un modelo de organización del cerebro, que se vuelve más preciso a medida que se añaden más datos. Además, para poder cuestionar el papel, la estructura y las relaciones entre las distintas partes del cerebro, los investigadores de Parietal han creado un lenguaje específico de este ámbito, Neurolang, que permite consultar conjuntos de datos para identificar automáticamente las estructuras cerebrales en una nueva imagen del cerebro. Este lenguaje tiene garantías formales y permite producir información probabilística con un grado limitado de certeza.



Conectividad funcional entre regiones cerebrales - © Inria _ PARIETAL

Nuevos modelos de aprendizaje humano

Los equipos de esta área estudian cómo las máquinas pueden adquirir modelos de conocimiento interactuando con su entorno, empujadas por mecanismos de curiosidad artificial (también llamada robótica de desarrollo). Se trata de un reto importante relacionado con la cuestión de la sostenibilidad de la IA, al aprender con un pequeño conjunto de ejemplos frente a los enormes conjuntos de datos que utilizan actualmente los sistemas de aprendizaje profundo, con las ya conocidas consecuencias en términos de recursos informáticos y consumo de energía.

FLOWERS

Robots y sistemas epigenéticos en movimiento

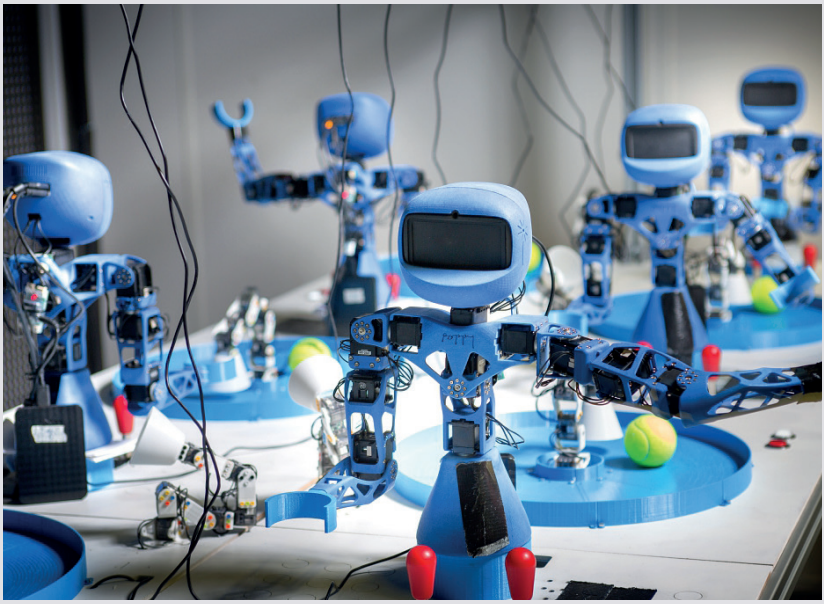
FLOWERS estudia modelos de desarrollo y aprendizaje abiertos. Estos modelos se utilizan como herramientas para ayudarnos a entender mejor cómo aprenden los niños, así como para construir máquinas de desarrollo que aprenden como estos últimos, con aplicaciones en robótica, interacción persona-computadora y tecnologías educativas.

Uno de los principales retos científicos de la inteligencia artificial y las ciencias cognitivas es comprender **cómo los seres humanos y las máquinas pueden adquirir eficazmente modelos del mundo, así como repertorios abiertos y acumulativos de habilidades a lo largo de un período prolongado.** Los procesos de desarrollo sensoriomotor, cognitivo y social se organizan a lo largo de fases ordenadas de complejidad creciente, y son el resultado de la compleja interacción entre el cerebro/cuerpo con su entorno físico y social.

Para avanzar en relación con la comprensión fundamental de los mecanismos del desarrollo, el equipo de FLOWERS ha desarrollado modelos computacionales que aprovechan técnicas avanzadas de aprendizaje automático, como el aprendizaje de **refuerzo profundo intrínsecamente motivado**, en estrecha colaboración con la psicología del desarrollo y la neurociencia. En particular, el equipo se ha centrado en modelos de aprendizaje y exploración intrínsecamente motivados (también llamados aprendizaje impulsado por la curiosidad), con mecanismos que permiten a los agentes aprender a representar y generar sus propios objetivos, auto-organizando un currículo de aprendizaje para un aprendizaje eficiente de los modelos del mundo y del repertorio de habilidades con recursos limitados de tiempo, energía y computación. El equipo también estudia cómo los mecanismos de aprendizaje autónomo pueden permitir a los humanos y a las máquinas adquirir **habilidades lingüísticas consolidadas, utilizando arquitecturas neuro-simbólicas para el aprendizaje de representaciones estructuradas y el manejo de la composibilidad y la generalización sistemáticas.**

Además de dar lugar a nuevas teorías y nuevos paradigmas experimentales para entender el desarrollo humano en la ciencia cognitiva, así

como a nuevos enfoques fundamentales para el **aprendizaje automático del desarrollo**, el equipo también ha explorado cómo estos modelos pueden encontrar aplicaciones en la robótica, interacción persona-computadora y tecnologías educativas. En el ámbito de la robótica, el equipo ha demostrado que la curiosidad artificial, combinada con el aprendizaje por imitación, puede proporcionar elementos esenciales que permitan a los robots adquirir múltiples tareas mediante la interacción natural con usuarios humanos inexpertos, por ejemplo en el contexto de la robótica asistencial. El equipo también demostró que los modelos de aprendizaje basado en la curiosidad **pueden transponerse en algoritmos para sistemas de tutoría inteligente**, permitiendo que el software educativo se adapte de forma incremental y dinámica a las particularidades de cada alumno humano, y proponiendo secuencias personalizadas de actividades de enseñanza. En cuanto a la interacción persona-computadora, el equipo ha demostrado cómo los algoritmos de aprendizaje incremental pueden utilizarse para eliminar la fase de calibración en determinadas interfaces cerebro-computadora.



Poppy Torso : aprendizaje impulsado por la curiosidad - © Inria _ Photo C. Morel

CoML

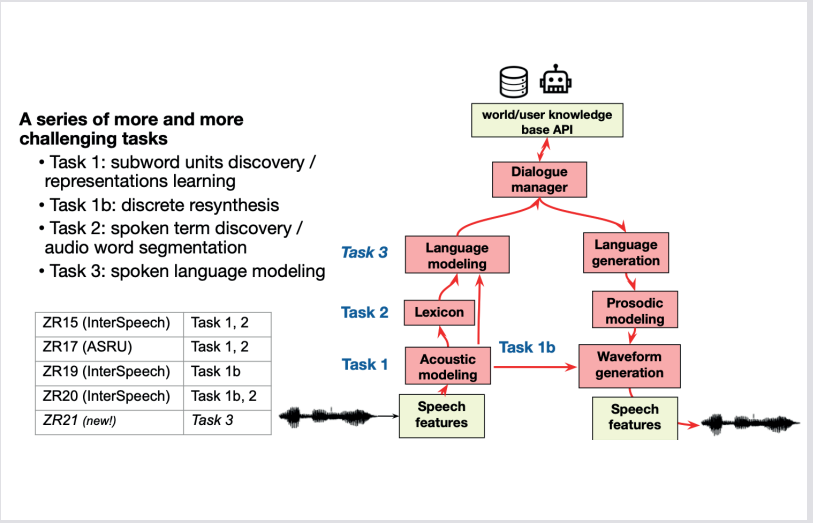
Aprendizaje automático cognitivo

El objetivo general de CoML es cerrar la brecha en relación con la flexibilidad cognitiva entre los humanos y el aprendizaje automático en el procesamiento del lenguaje y el razonamiento de sentido común, mediante la ingeniería inversa relativa a cómo los niños pequeños de entre 1 y 4 años de edad aprenden de su entorno. CoML trabaja sobre dos ejes: el primero, **IA del desarrollo**, la que se centra en la construcción de algoritmos de aprendizaje automático inspirados en los niños. El segundo eje, **Estudios cuantitativos del aprendizaje humano**, utiliza estos algoritmos para realizar análisis cuantitativos a gran escala del aprendizaje de los bebés humanos en la naturaleza en diversos entornos.

La IA del desarrollo se basa en la idea de que podría ser más sencillo construir una máquina que aprenda como un bebé, que construir una adulta (A. Turing, 1950). Las investigaciones sobre el desarrollo demuestran que los bebés aprenden de forma espontánea y autónoma el lenguaje, la cognición social y el sentido común, a partir de datos multimodales limitados y sin etiquetar y, en la mayoría de las culturas, sólo con una escasa supervisión directa de los adultos. Estudiamos cómo los algoritmos **auto-supervisados o débilmente supervisados** pueden descubrir representaciones o unidades discretas, como fonemas o palabras, a partir de la señal acústica bruta, sin ninguna etiqueta de experto (aprendizaje del habla con cero recursos). Exploramos los **sesgos inductivos** de los sistemas neuronales estudiando las condiciones de la aparición del lenguaje (aprendizaje del lenguaje con cero datos). Establecemos métricas y conjuntos de datos para sistemas no supervisados/auto-supervisados y reunimos **puntos de referencia o benchmarks y desafíos** para ayudar a construir una comunidad internacional en este ámbito general.

En los estudios cuantitativos del aprendizaje humano, analizamos las grabaciones naturalistas de larga duración de las interacciones entre los bebés y padres, para proporcionar **límites superiores e inferiores** a los datos que pueden producir un aprendizaje exitoso del lenguaje a través de la auto-supervisión o la supervisión débil (por ejemplo, un niño de 4 años requiere entre sólo 2.000 y 5.000 horas del habla dirigido para

aprender un sistema de diálogo de lenguaje hablado que funcione). Construimos **modelos causales** de crecimiento del lenguaje que predicen el vocabulario de los bebés en función de su input. También modelamos la adquisición de segundas lenguas en adultos. El equipo desarrolla una **plataforma** de hardware y software para ayudar a la recogida, anotación y análisis de datos a gran escala, preservando al mismo tiempo la privacidad y la seguridad (proyecto BeHive).



La serie Zero Resource Challenge: aprendizaje de representaciones del habla y el lenguaje mediante la autosupervisión a partir de audio sin procesar (www.zerospeech.com)

Acciones exploratorias (AEx)

AEx- ORIGINS - Basar la Inteligencia Artificial en los orígenes del comportamiento humano

Equipo-proyecto: FLOWERS

Uno de los objetivos más ambiciosos de la Inteligencia Artificial (IA) es la consecución de la llamada Inteligencia Artificial General (IAG), es decir, una IA que no se limite a la realización de un conjunto predefinido de tareas, sino que sea capaz de aplicar de manera general sus capacidades a cualquier tarea cognitiva que pueda ser resuelta por la inteligencia humana. Sin embargo, si bien la AGI está

fundamentalmente relacionada con las características de la inteligencia humana, la investigación en este campo rara vez considera los procesos que pueden haber guiado la aparición de capacidades cognitivas complejas durante la evolución de la especie. El AEx ORIGINS abordará esta laguna extrayendo principios computacionales de la literatura en Ecología del Comportamiento Humano y aplicándolos en la IA para mejorar la adquisición de comportamientos complejos en agentes artificiales.

AEx - ODiM - Herramientas informáticas de ayuda al diagnóstico de enfermedades mentales

Equipo-proyecto: SEMAGRAMME

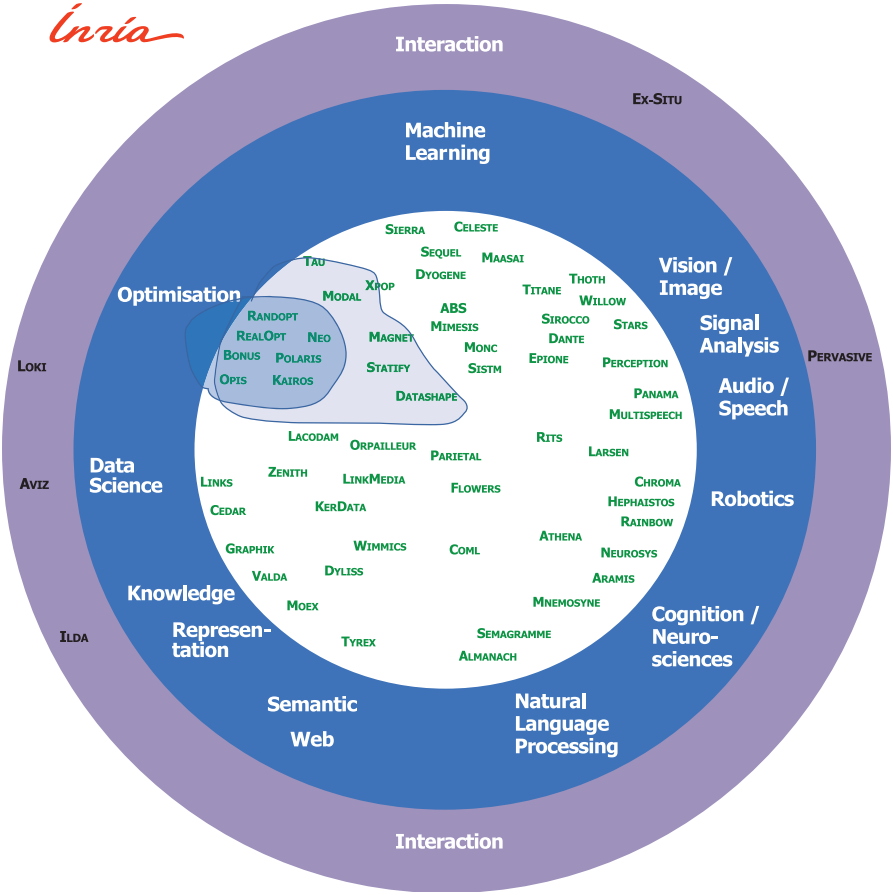
ODiM es un proyecto interdisciplinario en la interfaz entre la psiquiatría-psicopatología, lingüística, semántica formal y ciencias digitales. Su objetivo es desarrollar enfoques novedosos que ayuden al diagnóstico y detección de los trastornos psicóticos, ampliando los métodos de larga duración utilizados en psiquiatría. La producción de las herramientas está prevista para que un el mayor número posible de usuarios del sector de la salud mental (médicos, psiquiatras, psicólogos, logopedas, etc.) puedan utilizarlas.

Otros equipos de proyecto en este ámbito: NEUROSYS (Nancy)

5.8 Optimización

El cambio de siglo ha supuesto el desarrollo de la tecnología de optimización en la industria y en el campo científico correspondiente, en la frontera de la programación de restricciones (constraint programming), la programación matemática, la búsqueda local y el análisis numérico. En la actualidad, la tecnología de optimización ayuda al sector público, a las empresas y a las personas en cierta medida a tomar decisiones que utilicen mejor los recursos y se ajusten a requisitos específicos en un mundo cada vez más complejo. Efectivamente, la optimización y la toma de decisiones asistida por computador se están convirtiendo en una de las piedras angulares para ayudar a todo tipo de actividades humanas.

En un futuro más o menos cercano, se espera que la computación cuántica revolucione el campo de la optimización, permitiendo resolver problemas que hoy son intratables.



OPTIMIZACIÓN Y APRENDIZAJE AUTOMÁTICO

El aprendizaje automático se basa en la optimización numérica para el ajuste de los parámetros del modelo (miles de millones de ellos en el caso del aprendizaje profundo), por lo que desde hace décadas se han establecido estrechos vínculos entre ambos paradigmas. El uso del ML como componente de la optimización es una tendencia más reciente, en la que los modelos de aprendizaje automático–generalmente redes neuronales, gracias a sus propiedades de diferenciabilidad– permiten una optimización de extremo a extremo utilizando métodos de gradiente simples, siempre que se disponga de suficientes datos. Algunos retos se encuentran en la intersección de ambos enfoques.

Escalabilidad

Los modelos y los datos siguen creciendo exponencialmente a medida que aumenta el tamaño de los problemas. Es obligatorio diseñar métodos y algoritmos capaces de hacer frente a problemas cada vez más grandes sin utilizar recursos informáticos que aumenten exponencialmente. Esto vale para todo tipo de paradigmas de optimización, es decir, continuos, discretos o híbridos, y para todos los enfoques de aprendizaje automático.

Estructuras complejas

El aprendizaje automático y la optimización se ocupan de objetos complejos, es decir, no sólo de señales unidimensionales (sonido, imágenes, vídeos, etc.), sino también de estructuras como grafos, árboles, redes semánticas, etc. Aunque en muchos casos estas estructuras complejas pueden representarse mediante vectores gracias al desarrollo de modelos embebidos (embeddings) especializados, esto no es cierto para todas las estructuras, en particular, trabajar directamente con grafos puede ser especialmente útil, pero esto sigue siendo un reto.

Pruebas, confianza

Cuando se trata de aplicaciones del mundo real, todos los elementos que apoyan la confianza en los sistemas de IA/optimización utilizados son bienvenidos. Al principio de este capítulo, hemos abordado la cuestión genérica de la fiabilidad y confianza en la IA, en particular en el caso del aprendizaje automático.

Es necesario producir pruebas de convergencia o intervalos de confianza para los sistemas de optimización dentro de una cantidad razonable de recursos utilizados o tiempo de computación.

Uso adecuado de los modelos sustitutos

El primer uso histórico del aprendizaje automático dentro de un marco de optimización –todavía muy utilizado y exhaustivamente útil– ha sido el de proporcionar un modelo sustituto del sistema complejo en cuestión, que puede utilizarse de forma eficiente y fiel en lugar de ejecutar el sistema real, que en algunos casos ni siquiera es imaginable. El uso de estos modelos sustitutos implica desarrollar herramientas y métodos que ofrezcan garantías de que el modelo se aproxima lo suficiente a la realidad como para que los resultados puedan ponerse en práctica.

OPIS

Optimización de datos biomédicos a gran escala

OPIS es un nuevo proyecto de Inria-Saclay cuyo objetivo es abordar los retos que plantean los **métodos de optimización avanzada para el procesamiento de datos biomédicos a gran escala**. Los métodos de optimización están en el centro de muchos avances recientes en la inteligencia artificial, ya que una de las principales funcionalidades del cerebro es dar respuestas óptimas a los problemas que se nos plantean. OPIS busca métodos de optimización capaces de abordar datos con un gran tamaño de muestra ("N grande", por ejemplo, N=109) o muchas mediciones ("P grande", por ejemplo, P=104). Las metodologías que se explorarán se basarán en el análisis funcional no suavizado, la teoría del punto fijo o fixed point, las estrategias paralelas/distribuidas y las redes neuronales. Las nuevas herramientas de optimización que se desarrollarán se situarán en el marco general del procesamiento de señales gráficas, abarcando tanto los gráficos regulares (por ejemplo, las imágenes) como los no regulares (por ejemplo, las redes de regulación genética).

En concreto, OPIS trabaja en tres frentes:

1. Se diseñan nuevos algoritmos para resolver problemas de alta dimensión (que a veces implican hasta miles de millones de variables) que se

encuentran en problemas inversos, por ejemplo, la reconstrucción o restauración de imágenes, para aplicaciones médicas.

2. Se proponen nuevas estrategias para abordar problemas de minería de datos (data mining) formulados sobre grafos. Las estructuras de grafos permiten capturar las interacciones de sistemas complejos como los que existen en las redes biológicas.

3. Se investigan métodos de aprendizaje profundo poniendo énfasis en las garantías de robustez y en la capacidad de tener en cuenta la información previa.

Proponer mejores modelos de redes neuronales es de una importancia esencial en el contexto del diagnóstico o pronóstico de enfermedades a partir de imágenes médicas.

RANDOPT

Optimización aleatoria

El equipo RandOpt del centro de investigación de Inria en Saclay–Ile-de-France, que trabaja en conjunto con el CMAP de la École Polytechnique, se ocupa del análisis, desarrollo e implementación de métodos de optimización aleatorios de caja negra en el dominio continuo. RandOpt se centra en particular en los métodos de tipo CMA-ES y está interesado en la evaluación comparativa.

La especificidad de la optimización de caja negra es que los métodos están destinados a solucionar problemas caracterizados por la ausencia de propiedades: no lineales, no convexas y no suavizadas. Esto contrasta con la optimización basada en el gradiente y plantea, por un lado, algunos retos a la hora de desarrollar marcos teóricos, pero también obliga a complementar la teoría con investigaciones empíricas.

El objetivo último de RandOpt es proporcionar un software que sea útil para los profesionales. Se considera que la teoría es un medio para este

fin (más que un fin en sí mismo) y RandOpt cree firmemente que el ajuste de los parámetros forma parte de la tarea del diseñador.

Esto da forma, por un lado, a cuatro objetivos científicos principales:

1. **Desarrollar nuevos marcos teóricos** para orientar (a) el diseño de nuevos métodos de caja negra y (b) su análisis, permitiendo así

2. Proporcionar **pruebas fuera de las características clave** de los algoritmos adaptativos estocásticos, incluyendo el método de vanguardia CMA-ES: convergencia lineal y aprendizaje de información de segundo orden.

3. Desarrollar **algoritmos numéricos estocásticos de caja negra** siguiendo un **diseño de principios** en dominios con una fuerte necesidad práctica de métodos mucho mejores, a saber, **optimización restringida, multiobjetivo, a gran escala y costosa.** Implementar los métodos de forma que sean fáciles de usar. Y, por último, para

4. **Establecer nuevos estándares en la experimentación científica, la evaluación del rendimiento y la evaluación comparativa**, tanto para la optimización en espacios de búsqueda continuos o combinatorios. Esto debería permitir, en particular, avanzar en el estado de **reproducibilidad de los resultados de los trabajos científicos** en optimización.

OPTIMIZACIÓN Y RENDIMIENTO

En cuanto al diseño de técnicas eficaces de Inteligencia Artificial que se ocupen de tareas complejas y problemas de optimización, los principales retos son:

1. Obtener una comprensión más fundamental de qué es lo que hace que una tarea/problema sea difícil de resolver;

2. Acomodar la amplia gama de tareas/problemas complejos con respecto a la amplia gama de técnicas de resolución especializadas de una manera abstracta, flexible y eficiente;

3. La fertilización cruzada de los conocimientos de otras disciplinas, como

la HPC, la investigación operativa, etc., para aumentar la precisión y la eficacia;

4. Tratar con tareas/problemas de gran escala y costosos desde el punto de vista computacional; e

5. Incorporar la naturaleza multiobjetivo de muchas tareas/problemas prácticos, y escalar en supercomputadoras modernas.

BONUS

Gran optimización y computación a ultra escala

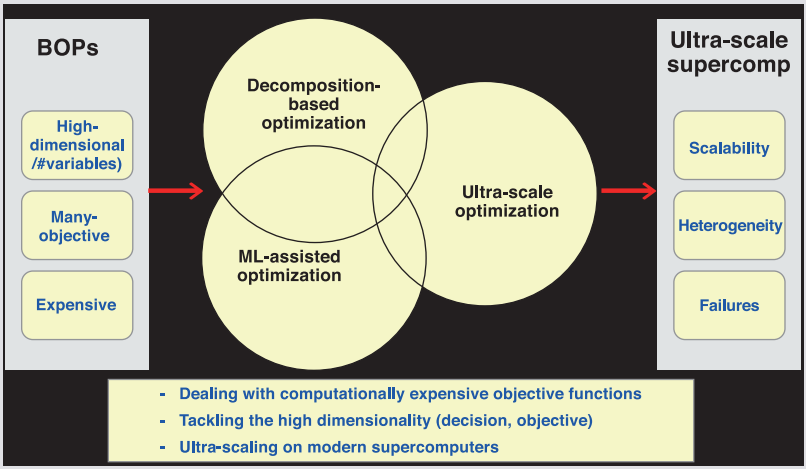
BONUS es un equipo de investigación conjunto entre Inria Lille – Nord Europe, CRIStAL (UMR 9189, Univ Lille, CNRS, EC Lille) y la Universidad de Lille. El equipo aborda grandes problemas de optimización, definidos por un gran número de parámetros, de variables de decisión o muchas funciones objetivo costosas en términos computacionales. Se centra en el diseño de técnicas de resolución eficaces a partir de la inteligencia computacional (búsqueda local estocástica, computación evolutiva) y la búsqueda combinatoria exacta (branch-and-bound) siguiendo tres líneas de investigación:

1. **Optimización basada en la descomposición:** Dada la escala particularmente voluminosa de los grandes problemas de optimización en cuanto a las variables y objetivos, BONUS desarrolla nuevas técnicas de descomposición dividiendo el problema objetivo original en sub-problemas más pequeños que son más fáciles de resolver y que están poco acoplados o son independientes. Resolver estos sub-problemas de forma simultánea y cooperativa es esencial para hacer frente a la "maldición de la dimensionalidad".

2. **Optimización asistida por aprendizaje automático:** Cuando se trata de problemas de alta dimensión y de objetivos procedentes de simulaciones u otros sistemas de caja negra, BONUS acopla técnicas de inteligencia computacional con metamodelos sustitutivos y otros algoritmos de aprendizaje automático para acelerar la convergencia del proceso de optimización y hacer frente a la naturaleza computacionalmente costosa de los grandes problemas de optimización.

3. **Optimización a ultra escala:** Con el fin de beneficiarse del paralelismo masivo que ofrecen las modernas supercomputadoras, BONUS se apoya en la computación a ultra escala para la resolución efectiva de grandes problemas de optimización, como el manejo de la gran cantidad de sub-problemas generados por la descomposición, o la evaluación paralela de objetivos y metamodelos basados en la simulación.

Desde el punto de vista del software, el objetivo de BONUS es integrar los enfoques que BONUS desarrollará en el marco de ParadisEO [3] (ParadisEO: http://paradiseo.gforge.inria.fr/) para permitir su reutilización dentro y fuera del equipo de Bonus. El mayor reto será ampliar ParadisEO para hacerlo más colaborativo con otros programas informáticos, incluyendo herramientas de aprendizaje automático, otros solucionadores (exactos) y simuladores.



BONUS colabora estrechamente con investigadores internacionales de la Universidad de Mons (Bélgica), la Universidad de Coimbra (Portugal), la Universidad de Shinshu (Japón), la City University (Hong Kong), la Universidad de Monash y la Universidad de Luxemburgo, en un esfuerzo por reflejar la fuerte sinergia entre la optimización, la inteligencia computacional y la computación paralela.

NEO

Ingeniería y operaciones de red

NEO se sitúa en la intersección de la Investigación Operativa y la Ciencia de las Redes

Los investigadores de NEO modelan situaciones que surgen en varios dominios de aplicación, que implican redes y sistemas distribuidos de una manera u otra, con el objetivo de tomar decisiones (posiblemente) óptimas utilizando las herramientas de la Investigación Operativa Estocástica. La IA moderna también se ocupa de las decisiones que toman (o sugieren) las máquinas basándose en algunos datos (aprendizaje automático). Por lo tanto, es natural que la IA distribuida se haya convertido en uno de los temas de investigación de NEO en los siguientes ejes:

1. Aprendizaje semi-supervisado en estructuras de grafos y sus implementaciones distribuidas.

2. Diseño de sistemas de aprendizaje automático distribuido a escala de Internet, tanto para el entrenamiento como para la inferencia, centrándose en el equilibrio (trade-off) entre el rendimiento y los costos económicos y medioambientales.

3. Modelos de aprendizaje multiagente basados en la teoría de juegos. Esto incluye la teoría de juegos evolutiva cuyo equilibrio consiste en puntos de reposo de la dinámica de tipo darwiniano, juegos dinámicos no cooperativos en los que la cooperación puede ser inducida mediante amenazas y castigos, y juegos de correspondencia que se han aplicado para las redes de recomendación.

4. Análisis de los límites fundamentales de la influencia de las políticas de provisión de información (sistemas de recomendación, medios de comunicación, redes sociales, etc.) en los tomadores de decisiones que participan en interacciones competitivas (mercados, sistemas de recursos compartidos, etc.).

El equipo colabora en estos temas con muchos socios industriales,

como Qwant, Nokia, Accenture, MyDataModels y Azursoft.

Otros temas de investigación relacionados con NEO son la asignación de recursos en las redes de comunicación, las redes sociales, la informática y las comunicaciones ecológicas y el desarrollo sostenible.

POLARIS

Análisis de rendimiento y optimización de grandes infraestructuras y sistemas

El objetivo del proyecto POLARIS es contribuir a la comprensión (desde la observación, el modelado y el análisis, hasta la optimización real mediante algoritmos adaptados) del rendimiento de los sistemas descentralizados a gran escala, como las supercomputadoras, infraestructuras en la nube, redes inalámbricas, redes inteligentes, los sistemas de transporte o incluso los sistemas de recomendación.

Una primera línea de investigación está dedicada al uso de técnicas de aprendizaje estadístico (inferencia bayesiana) para modelar el rendimiento esperado de los sistemas dispersos con el fin de construir visualizaciones de rendimiento agregadas, alimentar simuladores de dichos sistemas o detectar comportamientos anómalos.

En un contexto descentralizado, también es esencial diseñar sistemas que puedan adaptarse sin problemas a la carga de trabajo y al comportamiento evolutivo de sus componentes (usuarios, recursos, red). Obtener información fiel sobre la dinámica del sistema puede ser especialmente difícil, por lo que suele ser más eficiente diseñar sistemas que aprendan dinámicamente las mejores acciones a realizar mediante ensayo y error. Una característica clave del trabajo en el proyecto POLARIS es aprovechar regularmente el modelado teórico del juego para manejar situaciones en las que los recursos o la decisión se distribuyen entre varios agentes o incluso situaciones en las que un tomador de decisiones centralizado tiene que adaptarse a usuarios estratégicos.

Por ello, los miembros de POLARIS están especialmente interesados en

el diseño y análisis de algoritmos de aprendizaje adaptativo para sistemas multiagente, es decir, agentes que buscan mejorar progresivamente su rendimiento en una tarea específica (véase la figura). Los algoritmos resultantes no sólo deben aprender un equilibrio eficiente (Nash), sino que también deben ser capaces de hacerlo rápidamente (low regret), incluso cuando se enfrentan a las dificultades asociadas a un contexto descentralizado (falta de coordinación, mundo incierto, retraso en la información, retroalimentación limitada, etc.).

Una importante dirección de investigación en POLARIS se centra, por tanto, en el aprendizaje mediante refuerzo (Multi-armed bandits, Q-learning, online learning) y el aprendizaje activo en entornos con una o varias de las siguientes características:

- La retroalimentación es limitada (por ejemplo, no se dispone de gradientes o incluso de gradientes estocásticos, lo que obliga, por ejemplo, a recurrir a aproximaciones estocásticas);

- Entorno multiagente donde cada agente aprende, posiblemente, de una forma no sincronizada (es decir, las decisiones pueden tomarse de forma asíncrona, lo que plantea problemas de convergencia);

- Retroalimentación retardada (evitar las oscilaciones y cuantificar la degradación de la convergencia);

- Cargas de trabajo no estocásticas (por ejemplo, adversas) o no estacionarias (por ejemplo, en presencia de emergencias);

- Sistemas compuestos por un gran número de entidades, los que estudiamos mediante la aproximación de campo medio (juegos de campo medio y control de campo medio). Como efecto secundario, muchos de los conocimientos adquiridos pueden utilizarse a menudo para mejorar drásticamente la escalabilidad y el rendimiento de la implementación de técnicas de aprendizaje automático o profundo más estándar en las supercomputadoras.

KAIROS

Tiempo lógico multiforme para la concepción de sistemas cibernéticos

Las técnicas de aprendizaje automático (ML) (por ejemplo, las redes neuronales profundas) se han beneficiado de las plataformas de implementación eficientes (GPU y TPU) y de los métodos de compilación desarrollados por la comunidad de computación de alto rendimiento (HPC) para obtener viabilidad práctica y reconocimiento.

Mientras tanto, los sistemas embebidos de seguridad crítica (a menudo en tiempo real) se identifican como un lugar preferido para las aplicaciones de aprendizaje automático de la vida real (por ejemplo, conducción automatizada, modelos de gemelos digitales). Por lo tanto, resulta tentador y rentable combinar ambos dominios y, en particular, unificar:

1. Los métodos de compilación optimizados para las especificaciones de datos paralelos, desarrollados en la comunidad HPC/ML, y

2. Los métodos desarrollados en la comunidad integrada en tiempo real para proporcionar garantías de consumo de recursos en las situaciones más desfavorables para las especificaciones de tareas paralelas.

Basándose en la profunda proximidad entre los formalismos intermedios de los compiladores HPC/ML (MLIR/SSA) y los formalismos utilizados en el diseño en tiempo real (Lustre), el equipo de Kairos explora métodos para la especificación y la implementación (segura y eficiente) de aplicaciones incrustadas de alto rendimiento compatibles con el aprendizaje automático.

Otro equipo de proyecto en este ámbito: REALOPT (Burdeos)

5.9 IA e interacción humano-computador (HCI por sus siglas en inglés)

Los humanos pueden ahora delegar tareas como conducir un vehículo o pilotar un avión, y los sistemas de IA se proclaman regularmente como "mejores que los humanos" en diversas tareas de alto nivel. Sin embargo, los sistemas de IA no son perfectos y se ha mantenido o puesto a los humanos "en el bucle" de muchos sistemas críticos de seguridad basados en la IA para protegerlos de comportamientos inesperados del sistema. Desgraciadamente, esta disposición ha tenido consecuencias nefastas, como demuestran accidentes recientes como el de dos aviones comerciales Boeing 737 Max, en los que el sistema antibloqueo hizo que los aviones cayeran de morro veintiséis veces seguidas en menos de diez minutos sin entregar a los pilotos la información y el control necesarios para salvar el avión. Este tipo de accidentes son la consecuencia de una confianza ilimitada en la tecnología por encima de las habilidades humanas, y de un cambio de situaciones en las que los humanos delegan tareas pero siguen teniendo el control a aquellas en las que la computadora trata al humano como una fuente de información para un algoritmo. El "humano en el bucle" ***(human in the loop)*** es esencialmente un engranaje de la máquina, que asume la culpa cuando las cosas van mal. Estos sistemas no aprovechan al máximo el talento humano y las capacidades del sistema, sino que asumen que la computadora siempre puede calcular una solución óptima. ***Así pues, un reto importante tanto para la IA como para la HCI es crear una mejor división del trabajo entre los humanos y los computadores, aprovechando sus respectivos poderes y capacidades al tiempo que se reconocen sus limitaciones y debilidades.***

Otra vertiente que entrelaza la IA y la HCI está relacionada con las cantidades masivas de datos personales analizados por potentes algoritmos de aprendizaje automático. Nuestra interacción con el mundo digital se ha redefinido fundamentalmente: nuestras decisiones se controlan, se sujetan a influencias externas y a menudo se manipulan, lo que amenaza no sólo nuestra privacidad, sino también la democracia y los derechos humanos básicos. Asimismo, en este caso se ha cambiado el control humano sobre los procesos informáticos por el control informático sobre el comportamiento humano. *Un segundo reto importante es cómo aportar verdadera transparencia y*

capacidad de explicación a los sistemas de IA mediante interfaces de usuario y visualizaciones adecuadas.

Las aplicaciones actuales de las técnicas de IA en campos como el diagnóstico médico, las condenas judiciales o la conducción automatizada tienden a dejar de lado a los usuarios inexpertos: al automatizar tareas que antes realizaban los humanos, se puede mejorar la productividad para situaciones "normales". Pero las computadoras son extremadamente malas en cuanto a la gestión de casos excepcionales y es ilusorio pensar que una IA "mejor" cambiará significativamente esta situación. Los humanos, por otra parte, son muy buenos en el manejo de casos excepcionales, siempre que puedan mantenerse entrenados, pero son notoriamente malos en el seguimiento de actividades. *Un tercer reto importante es cómo combinar los sistemas interactivos y de IA para que cada uno aproveche los puntos fuertes del otro en el momento adecuado, minimizando al mismo tiempo las limitaciones de cada uno.*

Los sistemas modernos de IA se están volviendo tan complejos que los ingenieros necesitan nuevas herramientas simplemente para supervisar y gestionar su desarrollo, evolución, depuración y, en general, para entender lo que ocurre "bajo el capó" del sistema. Por ejemplo, los grandes entornos de aprendizaje automático vienen acompañados de sofisticadas herramientas para diseñarlos y programarlos[32]. La mayoría de los pasos de los sistemas de IA requieren herramientas para evaluar la calidad de los datos, las características, el entrenamiento y las decisiones; para entender el comportamiento de un sistema de IA en cualquier punto concreto; para supervisar y mejorar su calidad; para descubrir los sesgos y la incertidumbre en los resultados; y para entregar los resultados a los usuarios objetivo de forma significativa. *Un cuarto reto importante es crear herramientas mejores y más centradas en el usuario para los expertos que crean y evalúan los sistemas de IA.*

HCI para mejorar la IA

Además de las herramientas para mejorar la IA, la HCI debería ayudar a crear sistemas de IA más transparentes para que puedan ser evaluados por expertos en sus ámbitos de aplicación. Por ejemplo, la gestión de préstamos bancarios está cada vez más asistida por herramientas de IA y tiene un impacto directo en la vida de los ciudadanos. Algunas decisiones automatizadas han estado sujetas a sesgos estructurales difíciles de prever por los ingenieros de

32. K. Wongsuphasawat et al., "Visualizing Dataflow Graphs of Deep Learning Models in TensorFlow", en IEEE Transactions on Visualization and Computer Graphics, vol. 24, nº 1, pp. 1-12, enero de 2018.

IA, pero ciertamente detectables por los expertos en préstamos[33]. Sin embargo, para abordar estos sesgos se necesitan herramientas de comunicación entre los dos tipos de expertos para averiguar las causas y acordar las soluciones. En el caso de los préstamos, las causas se han encontrado en mediciones indirectas defectuosas utilizadas para evaluar a las personas y en datos de entrenamiento descompensados que representan erróneamente a las mujeres o a las minorías. Descubrir estos sesgos requiere un juicio humano, y éstos pueden ser de muy distinta índole.

La transparencia también se trata de algo más que meramente explicar las decisiones o mostrar la maquinaria, también consiste en explicar o tener en cuenta las capacidades de un sistema y sus limitaciones. Los vehículos auto-conducidos son buenos en algunas situaciones estándar, pero poco fiables en otras. Deben avisar al conductor para que recupere el control cuando sea necesario, lo que requiere que los sistemas de IA sean conscientes de su propio nivel o fiabilidad (algo que rara vez hacen) y que cedan el control a los humanos con elegancia, algo que es notoriamente difícil y requerirá más investigación.

Por último, los nuevos sistemas de aprendizaje automático tratan de aprender continuamente de los humanos mediante la interacción con ellos para completar sus conocimientos. Un sistema como el de búsqueda de Google mejora su precisión controlando el rango de los resultados que el usuario lee (en el que hace clic) tras una consulta de búsqueda. Este método sólo es eficaz para mejorar la "precisión" del buscador, pero no su recuerdo (si un resultado no se muestra, no puede ser clasificado). Encontrar métodos para aprender de forma interactiva y medir el aumento de la calidad y la usabilidad sigue siendo un problema complejo que requiere más investigación.

Aviz

Análisis y visualización

Aviz es un proyecto multidisciplinario que pretende mejorar la exploración y el análisis visual de grandes y complejos conjuntos de datos mediante una estrecha integración de los métodos de análisis con la visualización interactiva.

33. C. O'Neil, Weapons of Math Destruction, Crown Publishing, 2016.

Nuestro trabajo tiene el potencial de impactar a prácticamente todas las actividades humanas para y durante las cuales se recogen y gestionan datos que posteriormente hay que comprender. A menudo, las actividades relacionadas con los datos se caracterizan por el acceso a nuevos datos de los que tenemos poco o ningún conocimiento previo en relación con su estructura y contenido internos. En estos casos, necesitamos explorar primero los datos de forma interactiva para obtener información y, finalmente, poder actuar sobre el contenido de los datos. El análisis visual interactivo es especialmente útil en estos casos en los que los enfoques de análisis automático fallan y es necesario aprovechar y aumentar las capacidades humanas.

Dentro de este ámbito de investigación, Aviz se centra en cinco temas de investigación:

- Métodos para visualizar y navegar sin problemas a través de grandes conjuntos de datos;

- Métodos de análisis eficaces para reducir conjuntos de datos enormes a un tamaño de visualización aceptable;

- Interacción de visualización mediante nuevas capacidades y modalidades;

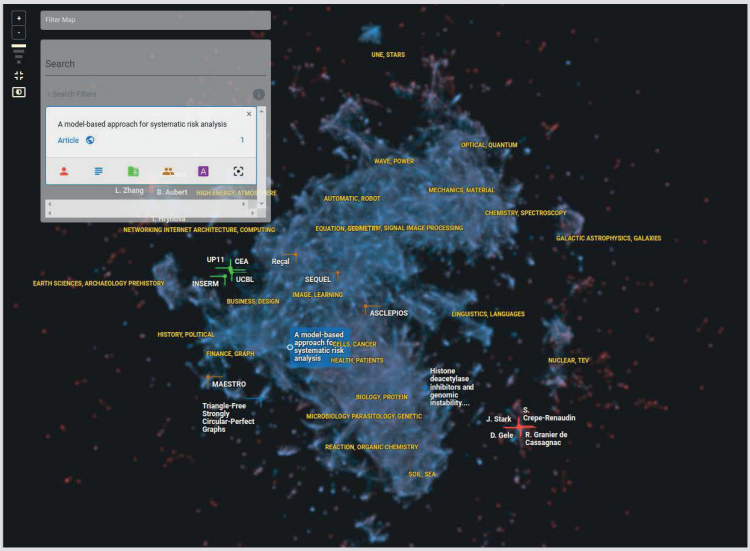
- Métodos de evaluación para valorar la eficacia de los métodos de visualización y análisis y la usabilidad de éstos;

- Herramientas de ingeniería para construir sistemas de análisis visual que puedan acceder, buscar, visualizar y analizar grandes conjuntos de datos con una respuesta fluida e interactiva.

En colaboración con el equipo-proyecto TAU, Aviz construye una visualización del repositorio HAL, el que contiene todas las publicaciones de las instituciones públicas de investigación francesas, utilizando proyecciones multidimensionales para crear un "mapa" resultante del análisis de Procesamiento del Lenguaje Natural (modelización de temas), la agrupación para recoger regiones temáticas sobre el mapa y encontrar etiquetas significativas. Todas estas técnicas relacionadas con la IA se reúnen mediante una interfaz de usuario basada en la Web para

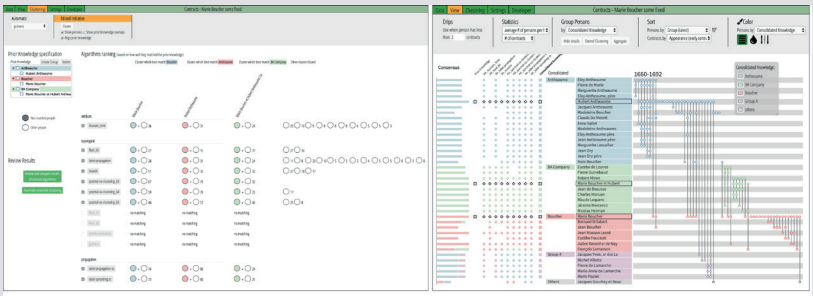
permitir a los investigadores de cualquier dominio explorar las publicaciones en torno a temas o autores, permitiendo que las complejas técnicas de IA sean exploradas por una gran audiencia de usuarios.

Véase [Philippe Caillou, Jonas Renault, JeanDaniel Fekete, Anne-Catherine Letournel, Michèle Sebag. Cartolabe: A Web-Based Scalable Visualization of Large Document Collections. IEEE CG&A 2020, pendiente de publicación] y https://cartolabe.fr



Cartolabe mostrando una visualización HAL, con 208984 autores (rojo) y 827156 artículos (azul).

Aviz también está trabajando en el análisis y la visualización de redes, para permitir que investigadores de redes –como historiadores, sociólogos o investigadores del cerebro– incorporen sus conocimientos previos a los métodos de agrupación (clustering) de conjuntos *[Alexis Pister, Paolo Buono, Jean-Daniel Fekete, Catherine Plaisant, Paola Valdivia. Integrating Prior Knowledge in Mixed Initiative Social Network Clustering. IEEE TVCG 2021, pendiente de publicación]*. Con PK-Clustering, los usuarios con poco conocimiento de los algoritmos de clustering pueden introducir parte de su conocimiento previo para seleccionar o dirigir mejor los algoritmos, en lugar de creer ciegamente en los resultados de un algoritmo en particular.



PK-Clustering, que muestra los resultados de nueve algoritmos de clustering mostrados como columnas de puntos a la izquierda (cada cluster o agrupación tienen un color), aplicados a la red a la derecha y consolidados en la columna más a la derecha contra el conocimiento previo.

La IA para mejorar la HCI

LOKI

Tecnología y conocimiento para la interacción

LOKI concibe a los computadores como herramientas que, en última instancia, podrían potenciar a las personas, centrándose en cómo pueden diseñarse y crearse dichas herramientas. Al comprender mejor los fenómenos que se producen en cada nivel de interacción y sus relaciones, reunimos los conocimientos y los ladrillos tecnológicos necesarios para conciliar la forma en que los sistemas interactivos se diseñan para, alrededor de y con las capacidades humanas. Nuestro ámbito de investigación abarca un amplio conjunto de entornos interactivos (computadoras de escritorio, dispositivos móviles, RV, ICC, etc.) y toma prestados sus métodos de campos tan variados como la psicología y neurociencia, la IA o el diseño y la ingeniería.

Dentro de nuestro objetivo de comprender mejor a los usuarios y diseñar sistemas que respondan adecuadamente a sus capacidades, recurrimos con frecuencia a las recientes aportaciones de la IA, sobre todo el aprendizaje automático y la optimización. Hemos desempeñado un papel decisivo en el diseño de la nueva norma de disposición del

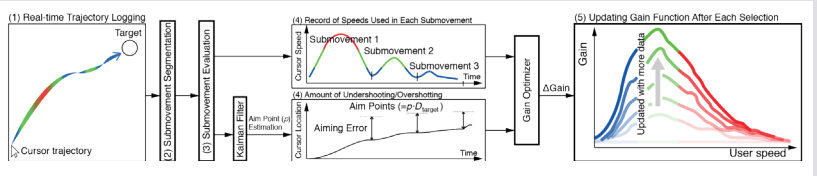
teclado francés [NF Z 71-300. http://norme-azerty.fr/] encargada por el Ministerio de Cultura de Francia, utilizando métodos de optimización combinatoria de última generación [A. Feit et al., Élaboration de la disposition AZERTY modernisée. 2018. https://hal.inria.fr/hal-01826476]. En colaboración con la Aalto University y el Max Planck Institute, desarrollamos un flujo de trabajo que permitía a expertos no técnicos en tipografía y lingüística iterar y evaluar ideas de diseño con un optimizador. Este optimizador, a su vez, era capaz de expresar las consecuencias de estas ideas en términos comprensibles de ergonomía y rendimiento tipográfico [A. Feit et al., AZERTY amélioré: Computational Design on a National Scale. En Communications of the ACM (en prensa)].

Con un enfoque diferente, los métodos de IA también pueden aprovecharse para adaptar dinámicamente las interfaces de usuario en función del perfil del usuario, el contexto de interacción o las necesidades. Por ejemplo, con colegas del University College de Londres, utilizamos métodos de agrupación jerárquica para adaptar el contenido mostrado en el perfil del usuario, en el contexto de la lectura de noticias en un dispositivo móvil [Constantinides et al., Exploring mobile news reading interactions for news app personalisation https://hal.inria.fr/hal-01252631]. Asimismo, tenemos previsto explorar el uso de métodos computacionales para anticiparse dinámicamente a las necesidades de los usuarios en el contexto de la interacción con el software enriquecido y ayudarles a descubrir características novedosas de las que aún no son conscientes (proyecto DISCOVERY de la ANR).

Muchos contextos interactivos podrían beneficiarse de una sinergia entre las aportaciones del usuario y la inteligencia del sistema. Una de nuestras hipótesis es que los usuarios son más propensos a aceptar una solución sugerida por una IA cuando han contribuido directamente al desarrollo de esa solución (por ejemplo, mediante aportaciones ocasionales explícitas), mientras que la IA proporciona una retroalimentación "honesta" que reconoce su posible imprecisión. Estamos explorando este tema en el marco del archivo de documentos manuscritos antiguos, que actualmente combina el escaneo de documentos con la transcripción manual o automática de forma secuencial. Siguiendo el mismo paradigma de colaboración entre humanos e inteligencia artificial, actualmente estamos explorando, junto con colegas de la University of Waterloo, cómo los usuarios confían en las palabras sugeridas por la

IA cuando escriben un texto. Investigamos cómo los usuarios gestionan la combinación entre escribir palabras con un teclado virtual y utilizar las sugerencias propuestas por la IA, en función de la precisión de las sugerencias y la eficacia de la interfaz. Esto ayudará a orientar el diseño de los sistemas interactivos, proporcionando formas de automatizar la tarea del usuario [Roy et al. Bajo revisión CHI 2021].

Interactuar con un sistema en tiempo real requiere la capacidad de reunir e interpretar flujos de datos continuos que pueden ser poco claros o carecer de semántica. La IA nos permite aprovechar mejor estas ricas señales y resolver problemas conocidos de la interfaz de forma novedosa y eficiente. La latencia, por ejemplo, sea perceptible o no [R. Jota et al., How Fast is Fast Enough? A Study of the Effects of Latency in Direct-touching Tasks. En Proc. of ACM CHI '13], es un castigo al rendimiento de la interacción. Hasta hace poco, su única solución era esperar a que el hardware mejore, lo que, sin embargo, va seguido inevitablemente de un software más exigente, que devuelve la latencia al punto de partida. Nosotros probamos un enfoque más independiente del hardware: aplicamos técnicas de optimización y estimación de última generación para afinar un algoritmo capaz de predecir con precisión los siguientes movimientos del cursor, lo cual utilizamos para compensar visualmente la latencia de extremo a extremo para el señalamiento relativo [M. Nancel et al., Next-Point Prediction for Direct Touch Using Finite-Time Derivative Estimation. En Proc. of ACM UIST '18. https://hal.inria.fr/hal-01893310]. Asimismo, utilizando algoritmos de optimización, y en colaboración con la University of Aalto y el KAIST, diseñamos una herramienta capaz de adaptar en tiempo real el perfil de aceleración de un cursor a las habilidades y hábitos de apuntar del usuario, ya sea controlado por un ratón, un trackpad, o incluso por gestos de la mano en el aire [B. Lee et al. AutoGain: Gain Function Adaptation with Submovement Efficiency Optimization. En Proc. ACM CHI '20. https://hal.inria.fr/hal-02918581].



Asociaciones entre humanos e inteligencia artificial

Algunos de los primeros pensadores, como J.C.R. Licklider y D. Engelbart, han propuesto el concepto de "simbiosis hombre-máquina"[34] o la visión de "aumentar el intelecto humano"[35], en la que los sistemas informáticos utilizan la IA para servir, y no para sustituir, la inteligencia y los conocimientos humanos. La creación de este tipo de ***asociaciones entre el ser humano y la IA*** es fundamental a la hora de combinar la IA y la HCI.

La interacción humano-computador se centra en la interacción entre el usuario y un sistema, que suponemos es una relación dinámica que cambia con el tiempo. Cuando tratamos con sistemas inteligentes, tanto el usuario como el sistema pueden tener influencia y control. Uno de los principales retos del diseño de la interacción es cómo gestionar este control compartido, idealmente dejando al usuario al mando del control de la interacción, pero al menos dándole un "consentimiento informado" sobre lo que está sucediendo. La perspectiva estándar del "humano en el bucle" trata al usuario humano como una entrada para alimentar el algoritmo y el éxito se define en términos de crear algoritmos más rápidos y de mayor rendimiento. Aunque crear mejores algoritmos sigue siendo un objetivo deseable, es fundamental que también adoptemos una perspectiva centrada en el ser humano, que defina el éxito en términos más cualitativos y orientados al usuario, lo que incluye un mayor rendimiento humano, pero también un aumento de las capacidades y la satisfacción de éste. Esta perspectiva también influye en la forma de ver los enfoques de iniciativa mixta. En lugar de intentar sustituir al usuario humano por un algoritmo, hacen hincapié en el papel continuo del primero dentro de la interacción. La mayor parte de la investigación actual sobre iniciativas mixtas sigue centrándose en el algoritmo, en lugar de mejorar las habilidades humanas. Las asociaciones entre humanos y la IA tratan de aprovechar las mejores características de los usuarios humanos y de los sistemas inteligentes, cuando la combinación supera lo que puede conseguir cualquiera de ellos por separado.

34. http://memex.org/licklider.pdf
35. https://www.dougengelbart.org/pubs/papers/scanned/Doug_EngelbartAugmentingHumanIntellect.pdf

ExSitu

Interacción en situaciones extremas

ExSitu explora los límites de la interacción: cómo los usuarios extremos interactúan con la tecnología en situaciones extremas. Nos interesan especialmente los profesionales creativos, artistas y diseñadores que reescriben las reglas al crear nuevas obras y los científicos que buscan comprender fenómenos complejos mediante la exploración creativa de grandes cantidades de datos. Estudiar hoy a estos usuarios avanzados no sólo nos ayudará a anticipar las tareas rutinarias del futuro, sino a avanzar en nuestra comprensión de la propia interacción.

En las prácticas creativas, el *aprendizaje automático centrado en el ser humano* facilita el flujo de trabajo para que los creativos exploren nuevas ideas y posibilidades. Hemos recopilado los recientes avances en investigación y desarrollo del aprendizaje automático centrado en el ser humano y la IA en las industrias creativas [B. Caramiaux et al. AI in the media and creative industries, New European Media (NEM), abril de 2019, pp. 1-35. https://hal.inria.fr/hal-02125504]. Asimismo, hemos explorado el uso de *Deep Reinforcement Learning* (aprendizaje de refuerzo profundo) en el contexto del diseño de sonido, comparando la exploración manual frente a la exploración por refuerzo. Demostramos que un explorador de sonido algorítmico que aprende de las preferencias humanas mejora el proceso creativo al permitir una exploración holística e integral frente a la exploración analítica que ofrecen las interfaces estándar.

También nos interesa diseñar *asociaciones eficaces entre humanos y computadores*, en las que los usuarios expertos controlan su interacción con la tecnología. En lugar de tratar a los usuarios humanos como la "entrada" o dato para alimentar un algoritmo informático, exploramos el aprendizaje automático centrado en el ser humano, cuyo objetivo es utilizar el aprendizaje automático y otras técnicas para aumentar las capacidades humanas. Nuestro objetivo específico es crear sistemas *co-adaptativos* que sean explorables, apropiables y elocuentes para el usuario. El proyecto CREATIV ERC Advanced desarrolló este enfoque y creó una serie de prototipos diseñados para aumentar el poder de expresión del usuario en los dispositivos móviles: CommandBoard [J.

Alvina et al. CommandBoard: Creating a General-Purpose Command Gesture Input Space for Soft Keyboards. Proc. UIST 2017. http://hal.inria.fr/hal-01679137]; FieldWard [J. Malloch et al. Fieldward y Pathward: Dynamic Guides for Defining Your Own. Proc. CHI 2017. http://hal.inria.fr/hal-01614267]; Expressive Keyboard [J. Alvina et al. Expressive Keyboards: Enriching Gesture-Typing on Mobile Devices. Proc. UIST 2016. http://hal.inria.fr/hal-01437054] .

Cuando trabajamos con profesionales creativos, no nos centramos en intentar que sean más creativos –ya lo son–, sino en proporcionarles herramientas que apoyen su propio proceso creativo personal. Entre esas herramientas está el uso de papel interactivo para ayudar a los compositores [Musink, Polyphony] y a los diseñadores [StickyLines, Enact]. También hemos explorado cómo los diseñadores y los sistemas inteligentes pueden compartir eficazmente la influencia o control según sus necesidades del momento con Semantic Collage [J. Koch et al. (2020) Semantic Collage. En Proc. DIS'20. https://dl.acm.org/doi/10.1145/3357236.3395494] e ImageSense: [J. Koch et al. (2020) ImageSense: An Intelligent Collaborative Ideation Tool to Support Diverse Human-Computer Partnerships. En Proc. ACM on Human Computer Interaction, Issue CSCW. https://hal.archives-ouvertes.fr/hal-02867303], junto con la Aalto University.

En el proyecto ***Bayesian Information Gain*** (BIG), en colaboración con Telecom París, utilizamos una técnica basada en el diseño experimental bayesiano, cuyo criterio es maximizar el concepto teórico de información mutua: en lugar de limitarse a interpretar las órdenes del usuario, BIG utiliza las entradas del usuario para actualizar sus conocimientos sobre el objetivo que pretende alcanzar y proporciona una salida que maximiza la ganancia de información esperada de la siguiente entrada. En otras palabras, el sistema desafía al usuario para que la interacción sea más eficiente. Hemos aplicado BIG a la navegación multiescalar [W. Liu et al. BIGnav: Bayesian Information Gain for Guiding Multiscale Navigation. Proc. CHI 2017. http://hal.inria.fr/hal-01677122] y a la recuperación de archivos [W. Liu et al. . BIGFile: Bayesian Information Gain for Fast File Retrieval. Proc. CHI 2018. http://hal.inria.fr/hal-01791754], con ganancias de rendimiento de hasta el 40% en comparación con las técnicas de navegación convencionales.

ILDA

Interacción con datos de gran tamaño

ILDA diseña sistemas interactivos centrados en los datos que proporcionan a los usuarios la información adecuada en el momento oportuno y les permiten manipular y compartir éstos de forma eficaz. Nuestro trabajo se centra en el diseño, el desarrollo y la evaluación de técnicas novedosas de interacción y visualización para empoderar a los usuarios en contextos tanto móviles como estáticos que implican una variedad de dispositivos de visualización, incluyendo: teléfonos inteligentes y tablets, auriculares de realidad aumentada, puestos de trabajo, escritorios, pantallas de tamaño de pared de ultra alta resolución. Nuestros temas de investigación incluyen nuevas formas de introducción y visualización tanto para grupos como para individuos, así como nuevas formas de interactuar con nuevos modelos de datos que permiten diversas estrategias de estructuración y consulta, dan una semántica procesable por la máquina a los datos y facilitan su interconexión. Investigamos formas de aprovechar esta riqueza desde la perspectiva de los usuarios, diseñando sistemas interactivos adaptados a las características específicas de los modelos de datos y la semántica de los datos, con un enfoque en los sistemas de importancia crítica y el análisis exploratorio de datos científicos.

Con colegas de París-Descartes y el equipo de ExSitu, investigamos las asociaciones entre humanos y la inteligencia artificial en el ámbito de la neurociencia y el análisis de series temporales (señales de EEG o electroencefalograma). Primero exploramos cómo ayudar a los neurocientíficos expertos a evaluar los patrones epileptiformes encontrados en las señales de EEG, combinando la visualización y el procesamiento automatizado en forma de algoritmos de búsqueda de similitud. Examinamos cómo el uso de diferentes visualizaciones puede afectar a la percepción de la similitud en las señales de EEG y cómo las diferentes visualizaciones pueden adaptarse mejor a las medidas de similitud [A.Gogolou, et al. Comparing Similarity Perception in Time Series Visualizations. IEEE TVCG 2019 (Proc InfoVis 2018), https://hal.inria.fr/hal-01845008]. Así, mostramos que la noción de similitud depende de la visualización y la necesidad de hacer coincidir los procesos automatizados con las representaciones

visuales adecuadas. Otros trabajos también ayudan a los expertos a consultar colecciones masivas de series de datos (como las bases de datos de EEG) en tiempo real. Nosotros aportamos resultados de búsqueda de similitud progresiva en grandes colecciones de series temporales (100 GB) y mostramos cómo estos pueden reducir los tiempos de espera para los usuarios, ya que observamos que se encuentran respuestas aproximadas de alta calidad muy pronto, por ejemplo, en menos de un segundo [A.Gogolou et al. Progressive Similarity Search on Time Series Data. Proc BigVis 2019, https://hal.inria.fr/hal-02103998v1]. Sin embargo, es importante que los usuarios puedan determinar la calidad de estas primeras respuestas y decidir si necesitan esperar más para obtener mejores coincidencias. Para ello, hemos trabajado en proporcionar límites probabilísticos de distancia y error, para ayudar a los analistas a evaluar la calidad de sus resultados progresivos [A.Gogolou et al. Data Series Progressive Similarity Search with Probabilistic Quality Guarantees. Proc ACM SIGMOD 2020, https://hal.inria.fr/hal-02560760v1].

También colaboramos desde hace tiempo con colegas de INRAe, donde combinamos la exploración visual con la computación evolutiva para ayudar a guiar a los expertos en la exploración de grandes conjuntos de datos multidimensionales. Nuestro framework (Evolutionary Visual Exploration – EVE), utiliza un algoritmo evolutivo interactivo para dirigir la exploración de conjuntos de datos multidimensionales hacia proyecciones bidimensionales que son de interés para el analista [N.Boukhelifa et al. Evolutionary Visual Exploration: Evaluation of an IEC Framework for Guided Visual Search Evolutionary Computation, En Evolutionary Computation, MIT Press, 2018]. Nuestro método combina sin problemas las métricas calculadas automáticamente y las entradas del usuario para proponerle vistas pertinentes. Este trabajo ha dado lugar a un prototipo de aplicación que ha sido utilizado por expertos de dominio en diferentes campos para formular hipótesis interesantes y alcanzar nuevos conocimientos al explorar libremente [N.Boukhelifa et al. Evolutionary Visual Exploration: Evaluation with Expert Users. En Computer Graphics Forum 2013, https://hal.inria.fr/hal02005699v1]; ha actuado como plataforma colaborativa para que equipos de investigadores exploren intercambios de información (trade-offs) [N.Boukhelifa et al. An Exploratory Study on Visual Exploration of Model Simulations by Multiple Types of Experts. Proc ACM CHI 2019, https://hal.inria.fr/hal-02005699v1]; y ha

dado lugar a investigaciones sobre la mejor manera de probar y evaluar marcos como EVE, que incorporan inteligencia humana y artificial que trabajan juntas para tomar decisiones.

Signal+IA como resultado o entrada que alimenta la HCI

Los sistemas interactivos aprovechan cada vez más los sensores que captan la rica información del usuario, como la voz, la mirada, los gestos o la actividad cerebral. La HCI utiliza técnicas de IA, en particular el aprendizaje automático, para analizar, reconocer o clasificar estas señales. El contexto de la interacción crea restricciones específicas que ponen a prueba los límites de las actuales técnicas de IA: el procesamiento debe producirse en tiempo real, a la escala del bucle de percepción-acción humana (normalmente menos de 100 ms y a veces mucho menos); los modelos a menudo deben entrenarse con muy pocas muestras, por ejemplo, un usuario sólo está dispuesto a mostrar un gesto una o dos veces y esperar que el sistema lo reconozca sólidamente a partir de entonces; el modelo debe adaptarse a los cambios en el comportamiento del usuario a lo largo del tiempo. En muchos casos, el reconocimiento debe producirse de forma progresiva, a medida que llega la señal, de modo que el sistema pueda proporcionar una respuesta y una solicitud ***(feed-forward)*** en tiempo real, como ejemplifica la guía dinámica Octopocus para la entrada de gestos[36]. Además, la captación continua de datos, por ejemplo, los datos de movimiento de un sensor Kinect, deben segmentarse en tiempo real, además de los segmentos que se reconocen. El aprendizaje automático interactivo, el aprendizaje por refuerzo, el aprendizaje activo y el aprendizaje en línea proporcionan enfoques potenciales para abordar estos problemas.

PERVASIVE

El proyecto de Inria PERVASIVE INTERACTION (interacción ubicua) desarrolla teorías y modelos para una interacción sociable y consciente del contexto con sistemas y servicios compuestos por objetos ordinarios a los que se les ha añadido la capacidad de percibir, actuar, comunicarse e interactuar con los seres humanos y con el entorno (objetos inteligentes). La capacidad de interconectar objetos inteligentes hace posible el montaje de nuevas formas de sistemas y servicios en entornos humanos ordinarios.

La interacción ubicua explora el uso de modelos de situación como base para el comportamiento circunstancial de los objetos inteligentes. La investigación se basa en experimentos de interacción en entornos variables con personas, ambientes y la computación omnipresente.

El programa de investigación aborda la siguiente cuestión: ¿puede el modelado de situaciones proporcionar una teoría para el comportamiento contextualizado de los objetos inteligentes? El programa está impulsado por las siguientes cuatro preguntas de investigación:

P1: ¿Cuáles son las técnicas computacionales más adecuadas para adquirir y utilizar modelos de situación para el comportamiento contextualizado de los objetos inteligentes?

P2: ¿Qué técnicas de percepción y acción son las más apropiadas para los objetos inteligentes conscientes del contexto?

P3: ¿Podemos utilizar el modelado de situaciones como base para una interacción social con objetos inteligentes?

P4: ¿Podemos utilizar los objetos inteligentes contextualizados como una forma de comunicación inmersiva?

Lo anterior se organiza en cuatro áreas de investigación que interactúan entre sí y que responden a estas preguntas de investigación:

36. O.Bau & W. Mackay. OctoPocus: A Dynamic Guide for Learning Gesture-Based Command Sets. UIST 2008. http://dl.acm.org/citation.cfm?id=1449724]

RA1. Adquisición y uso de modelos de situación (Q1)

RA2. Percepción de personas, actividades y emociones (Q2)

RA3. Interacción sociable con los humanos (Q3)

RA4. Interacción con objetos inteligentes ubicuos (Q4)

IA explicativa

La IA explicativa suele caracterizarse por informar a los usuarios cómo funciona un algoritmo. Sin embargo, una verdadera perspectiva de interacción humano-computadora desplaza el enfoque, a la luz del argumento de que los usuarios rara vez se preocupan por los detalles del funcionamiento del algoritmo y, en cambio, se preocupan más por la forma en que dichos algoritmos pueden afectarlos personalmente, así como por su capacidad para realizar la tarea en cuestión. Por tanto, el principal reto de la IA explicativa centrada en el usuario es cómo proporcionarle la información en términos que éste entienda. Los usuarios deben ser capaces de visualizar cómo el sistema de IA está interpretando y reaccionando a su comportamiento, así como qué decisiones está tomando y por qué. Los usuarios deben poder intervenir en el proceso, no sólo para descubrir cómo y por qué la IA ha realizado una determinada interacción, sino también para disponer de formas sencillas de informar a la IA cuando esas decisiones son incorrectas y sugerir soluciones mejores. Sistemas como Fieldward y Pathwar[37] proporcionan tanto una retroalimentación visual como una reacción progresiva a medida que el usuario dibuja un nuevo comando gestual propuesto. La IA interpreta dinámicamente el gesto a medida que se dibuja y proporciona una clasificación continua que se revela a través de un mapa de calor de colores cambiantes o de continuaciones de gestos. Esto muestra al usuario cómo la IA ha interpretado el gesto en ese instante y sugiere estrategias alternativas para generar con éxito un nuevo y único comando.

Sesgos cognitivos, ética y cuestiones jurídicas

La equidad, la explicabilidad y la responsabilidad son propiedades críticas para la aceptabilidad de los sistemas de IA en una amplia gama de dominios. Sin embargo, estas propiedades deben evaluarse desde una perspectiva humana y no sólo desde la perspectiva del sistema. Por ejemplo, los experimentos seminales de Tversky y Kahneman en economía conductual muestran que la percepción humana de la justicia no siempre es racional y depende en gran medida de la información contextual, como por ejemplo la forma en que se formula la pregunta. En general, se sabe que muchos sesgos cognitivos afectan a la toma de decisiones y al razonamiento humano, como el sesgo de confirmación y de anclaje. Esto implica que debemos adoptar métodos experimentales centrados en la HCI que incluyan a los participantes, en lugar de basarnos únicamente en las simulaciones y mediciones habituales en la investigación de la IA. Sin embargo, esto también plantea cuestiones éticas sobre si los sistemas de IA deben tener en cuenta los sesgos humanos y cómo hacerlo, ya sea reproduciéndolos o, por el contrario, combatiéndolos.

Otro tipo de sesgo tiene que ver con los conjuntos de entrenamiento de los sistemas inteligentes. Estudios recientes han demostrado que los algoritmos de detección de rostros son extremadamente precisos para los hombres blancos (más del 98%), menos precisos para las mujeres blancas y menos del 30% para las mujeres negras. Cuando los ingenieros jóvenes, hombres blancos, seleccionan conjuntos de entrenamiento de personas que se parecen a ellos, el resultado estará sesgado cuando se aplique a la población general, como cuando se utilicen estos datos para identificar posibles delincuentes o posibles candidatos a un puesto de trabajo.

La delegación de tareas y decisiones en los sistemas de IA plantea otras cuestiones éticas y jurídicas, en particular sobre la rendición de cuentas y la responsabilidad. Aunque parece haber consenso en que los seres humanos deben ser responsables en última instancia de las decisiones tomadas por los sistemas de IA, la tentación es culpar al usuario en lugar de al diseñador del sistema, como se ejemplificó en el accidente en el que murió el conductor de un coche autónomo. Una cuestión clave en este caso es si la interfaz del sistema de IA proporcionó al usuario información suficiente para evitar el accidente y si tuvo en cuenta los rasgos y el comportamiento humanos. Asumir que los

37. J. Malloch et al. Fieldward and Pathward: Dynamic Guides for Defining Your Own. Proc. CHI 2017. http://hal.inria.fr/hal-01614267]

usuarios siempre permanecerán en un estado de alerta elevado después de horas de conducción sin accidentes es una decisión de diseño fundamentalmente pobre, no un fallo del usuario humano. Las cuestiones éticas deben abordarse dentro del entorno socio-técnico más amplio en el que opera el sistema.

6

Colaboración europea e internacional en materia de IA en Inria

COLABORACIONES EN IA: LA VISIÓN DE INRIA

Las acciones de cooperación europea e internacional de Inria tienen como objetivo promover el intercambio entre Inria y las zonas geográficas más dinámicas, al tiempo que defienden los valores europeos para una IA centrada en el ser humano[38]. El contexto es bien conocido: la carrera por la inversión en determinadas zonas del mundo, el papel de China y Estados Unidos en la IA y la competencia por el talento de prestigiosas instituciones académicas extranjeras y de actores privados en la IA.

Este contexto anima al instituto a reforzar las colaboraciones susceptibles de impulsar la calidad del trabajo de Inria, garantizar la visibilidad y el posicionamiento de los equipos al mejor nivel europeo e internacional, pero también enriquecer el debate del instituto sobre el impacto de la IA en nuestras sociedades.

Además de los vínculos que se establecen de forma natural entre los investigadores a través de colaboraciones e intercambios informales, Inria, como instituto público nacional de tecnología digital, construye su política internacional a través de acuerdos específicos con socios, teniendo en cuenta las directrices de la estrategia internacional de Francia, los obstáculos específicos a los que se enfrenta y el marco europeo.

CONTRIBUCIÓN A LOS ESFUERZOS EUROPEOS DE I+I EN MATERIA DE IA

Los puntos fuertes de Europa son la calidad de sus investigadores e ingenieros, su formación y sus aplicaciones. Consciente de los retos de la soberanía, la UE ha adoptado una estrategia centrada en el ser humano, abogando por principios éticos[39]. La implicación de Inria en los esfuerzos europeos en materia de IA se basa en tres dimensiones: la integración en redes, la participación en proyectos a gran escala y una sólida contribución a la investigación exploratoria, especialmente a través de proyectos financiados por el ERC.

(I) INTEGRACIÓN EN REDES

Inria es miembro de BDVA (Big Data Value Association) y EU Robotics, que son asociaciones europeas que reúnen a socios industriales y académicos activos en los campos de los datos y la robótica, coordinando respectivamente

38. The Ethics Guidelines for Trustworthy Artificial Intelligence (AI), AI HLEG, abril de 2019 T
39. White Paper on Artificial Intelligence: a European approach to excellence and trust, CE, febrero de 2020.

las correspondientes Asociaciones Público-Privadas (APP). Además, INRIA participa en la propuesta de APP de IA/Datos/Robótica que se ha presentado a la Comisión Europea en 2021.

Además, han surgido en Europa varias redes de orientación académica que incluyen:

- Redes por iniciativa de comunidades científicas, como CLAIRE (Confederation of Artificial Intelligence Research Laboratories in Europe) y ELLIS (European Laboratory for Learning and Intelligent Systems). Inria apoya institucionalmente la iniciativa CLAIRE, pero también reconoce el apoyo a la iniciativa ELLIS de una parte de sus investigadores;

- Redes por iniciativa de la Comisión Europea para ayudar a estructurar las distintas comunidades de IA y estimular el diálogo y la convergencia entre ellas.

En cuanto a la red apoyada por la Comisión Europea a través del programa Horizon 2020, Inria participa en tres proyectos que se iniciaron el 1 de septiembre de 2020: los proyectos TAILOR y HumanAI R&I y la acción de apoyo y coordinación VISION. Estos proyectos sientan las bases de un ecosistema europeo de investigación e innovación de categoría mundial, para implantar una IA segura y fiable que respete los valores defendidos por la Unión Europea. Algunos investigadores de Inria son también integrantes del proyecto ELISE.

TAILOR pretende reforzar los vínculos entre los actores de la investigación académica, pública e industrial para desarrollar la base científica de la IA de confianza. Para ello, combina el aprendizaje, la optimización y el razonamiento para producir sistemas de IA que garanticen los requisitos de fiabilidad, seguridad, transparencia y respeto de las actividades humanas, y optimicen los beneficios esperados reduciendo los posibles daños.

HumanAINet pretende desarrollar una IA que sea segura, fiable y capaz de adaptarse a entornos reales e interactuar adecuadamente en contextos sociales complejos. El objetivo es promover sistemas de IA que mejoren las capacidades humanas y proporcionen apoyo a los individuos y a la sociedad en su conjunto, respetando al mismo tiempo la autonomía y la autodeterminación humanas.

ELISE reúne la mejor investigación europea en aprendizaje automático para crear una red de inteligencia artificial. Si bien ELISE parte del aprendizaje

automático como la tecnología central actual de la IA, la red invita a todas las formas de pensamiento, considerando todo tipo de datos, aplicables a casi todos los sectores de la ciencia y la industria.

VISION pretende coordinar la actividad de las cuatro redes europeas de excelencia en materia de IA (TAILOR, HumanAINet, ELISE y AI4Media), para contribuir a posicionar la investigación europea como protagonista de la IA. Para ello es necesario superar la fragmentación de la comunidad de la IA en Europa y estimular las sinergias para que surja la próxima generación de herramientas y sistemas de IA fiables, basados en métodos que abarquen una gama más amplia de técnicas de IA.

(II) COLABORACIONES A TRAVÉS DE GRANDES PROYECTOS DE INVESTIGACIÓN

Los proyectos a gran escala complementan y amplían el trabajo realizado en Inria.

AI4EU es el proyecto que tiene como objetivo construir la plataforma europea a medida ***(on demand),*** que es poner la tecnología de la IA al servicio de todos y, como tal, reducir las barreras a la innovación, estimular la transferencia de tecnología y facilitar el crecimiento de las nuevas empresas y las PYME en todos los sectores económicos.

TRUST-AI y ALMA son dos proyectos de investigación fundamentales que pretenden hacer avanzar la IA centrada en el ser humano. Más concretamente, TRUST-AI pretende integrar la noción de explicabilidad en la fase de aprendizaje de los modelos de "caja negra", sin comprometer su rendimiento. ALMA se basa en el paradigma del aprendizaje automático algebraico (AML por sus siglas en inglés), que produce modelos generalizadores a partir de la integración semántica de los datos en estructuras algebraicas discretas, lo que presenta una serie de ventajas sobre los modelos de aprendizaje estadístico.

(III) EXCELENCIA CIENTÍFICA PROMOVIDA POR EL CEI

Desde el lanzamiento del CEI (Consejo Europeo de Investigación) en 2007, Inria ha obtenido 59 subvenciones individuales (Starting, Consolidator, Advanced), 2 subvenciones Synergy y 9 subvenciones Proof of Concept (PoC). En el campo de la IA, Inria cuenta con 17 galardonados por el CEI, uno de los cuales obtuvo una financiación PoC además de su subvención individual *(véase la tabla siguiente y la lista en el anexo).*

Aprendizaje automático y sus aplicaciones	*Francis Bach, Julien Mairal, Alessandro Rudi, George Drettakis (aplicación)*
Visión por computador y procesamiento señal-imagen	*Cordelia Schmid, Ivan Laptev, Josef Sivic, Jean Ponce, Rémi Gribonval, Radu Horaud, Alexandre Gramfort, Emilie Chouzenoux*
Imágenes médicas	*Nicolas Ayache, Stanley Durrleman, Rachid Deriche*
Robótica	*Pierre-Yves Oudeyer, Jean-Baptiste Mouret*

ASOCIACIONES INTERNACIONALES DE INRIA EN MATERIA DE IA

Desde el año 2017, observamos un aumento de las políticas públicas y las estrategias nacionales en relación con la IA que son promovidas por las autoridades nacionales y que a menudo incluyen una dimensión internacional. Esto da lugar a múltiples requisitos. Así, estos contactos pueden generar acuerdos para explorar las oportunidades y los retos de la colaboración, en un enfoque de arriba hacia abajo.

Por ejemplo, a través de Inria Chile[40], el instituto participa en acciones y proyectos en el ámbito de la IA o sus aplicaciones. Inria Chile, en colaboración con instituciones locales, contribuye a la definición de la política chilena de IA llevada a cabo por el Ministerio de Ciencia, Tecnología, Conocimiento e Innovación y el Senado.

Además, Inria apoya las colaboraciones internacionales, en un enfoque ascendente, gracias a los incentivos ad hoc (Inria International Labs, Equipos Asociados, programas de movilidad), que permiten a Inria seguir atendiendo a las oportunidades de cooperación.

Por último, dado que los avances de la IA proceden en gran medida del sector privado, Inria opta a veces por establecer colaboraciones con actores industriales internacionales con importantes capacidades de I+D (véase el programa de investigación a largo plazo de Inria y Fujitsu sobre IA y procesamiento de ***big data***).

40. https://www.inria.fr/fr/centre-inria-chile

Además de esta política de vigilancia internacional, Inria centra actualmente sus esfuerzos de colaboración en el campo de la IA en tres áreas geográficas: Europa bilateral, Asia y Norteamérica.

EUROPA BILATERAL

Asociación Inria-DFKI

Tras el Tratado de Aquisgrán de 22 de enero de 2019 firmado entre Alemania y Francia para promover los esfuerzos conjuntos en el ámbito de la IA, Inria y DFKI celebraron en enero de 2020 un memorando de entendimiento en el que se comprometen a poner en marcha un programa conjunto de investigación e innovación. Este programa abarca los ámbitos de la IA para la industria 4.0, la IA para las tecnologías portátiles, la IA y ciberseguridad, y la cooperación entre humanos y robots. El Memorando de Entendimiento también forma parte de un compromiso conjunto dentro de la red CLAIRE.

Asociación entre Inria y University College of London

Firmado a finales de 2019, el acuerdo entre Inria y University College of London (UCL) formaliza la colaboración entre Inria y la UCL. Esta colaboración está diseñada para crecer y ampliarse para así incluir a otros socios londinenses.

ASIA

Dos países se consideran ahora prioritarios para el Instituto a la hora de establecer una cooperación en materia de inteligencia artificial en Asia: Japón y Singapur.

Japón

Existen muchas similitudes entre las visiones japonesa y francesa (y europea) de la IA: el enfoque japonés de la "IA centrada en el ser humano" se hace eco del concepto de IA para la humanidad de la estrategia francesa, y se considera que el intercambio seguro de datos y recursos entre socios de confianza permite ganar competitividad.

Además, en ambas estrategias nacionales se identifican los sectores de la movilidad y la salud como sectores prioritarios para la aplicación de la IA. Por último, los dos países también convergen en el uso de la IA para mejorar la

productividad, la consideración de los aspectos medioambientales y la necesidad de formar más talento en este campo.

En junio de 2019, Inria firmó un Memorando de Entendimiento de cuatro años con el Departamento de Tecnología de la Información y Factores Humanos del National Institute for Advanced Science and Industrial Science and Technology -AIST, que reúne ocho centros de investigación, incluido el Artificial Intelligence Research Centre (AIRC). Este convenio pretende reforzar la cooperación Inria-AIST, especialmente en el campo de la IA y la robótica, mediante el desarrollo de intercambios científicos y proyectos de investigación conjuntos.

Singapur

En 2018 se firmó un acuerdo de cooperación entre la National University of Singapore (NUS), como operadora del plan AI Singapore, e Inria, el CNRS y el INSERM. Este acuerdo tiene como objetivo promover el desarrollo de actividades conjuntas en IA y tecnologías digitales inteligentes, en las áreas de cooperación en IA y Salud; IA explicativa; aprendizaje federado o colaborativo; procesamiento automático del lenguaje natural; y confidencialidad, seguridad y responsabilidad en el intercambio de datos.

NORTEAMÉRICA

Tras una larga cooperación entre los equipos de proyectos de Inria y los investigadores norteamericanos en el campo de la IA, el Instituto lleva varios años formalizando asociaciones con actores muy visibles en la escena internacional e investigadores de renombre en este campo, principalmente en el ámbito de los métodos y herramientas fundamentales para el aprendizaje y el análisis de datos.

Estados Unidos de América

El Centre for Data Science and Courant Institute of Mathematical Sciences colabora activamente en el acuerdo entre New York University e Inria firmado en mayo de 2017, el que tiene una duración de 5 años. El programa conjunto ha permitido financiar proyectos de colaboración y visitas de investigadores y estudiantes de doctorado, así como la estancia de larga duración de un investigador senior de Inria (Jean Ponce).

Canadá

Inria y el CIFAR (Canadian Institute for Advanced Research) firmaron un acuerdo en enero de 2015, el que se está renovando actualmente. Inria participa en el programa "Neural Computing and Adaptive Perception", ahora llamado "Machine Learning, Biological Learning". Este programa está coordinado por Yann Le Cun (NYU y Facebook) y Yoshua Bengio (Université de Montréal). Los equipos de los proyectos WILLOW y SIERRA participan en las actividades de este grupo. Su principal objetivo es comprender los principios que subyacen a la inteligencia natural y artificial, así como dilucidar los mecanismos por los que el aprendizaje puede conducir a la aparición de la inteligencia.

Además de estas dos asociaciones, se apoyan a cinco colaboraciones en el marco del programa de Equipos Asociados de Inria:

- Carnegie Mellon University (Equipo Asociado de GAYA sobre modelos semánticos y geométricos para la interpretación de vídeos);

- University of Southern California (Equipo Asociado de LEGO sobre procesamiento automático del lenguaje);

- Stanford University (Equipo Asociado de Meta&Co sobre aprendizaje automático y procesamiento automático del lenguaje para el meta-análisis de asociaciones neuro-cognitivas) y el Equipo Asociado de Geomstat sobre anatomía algorítmica – aplicación de métodos de aprendizaje en neurociencia); y

- El Argonne National Laboratory (Equipo Asociado de UNIFY sobre aspectos de la IA como complemento para optimizar los flujos de trabajo híbridos que combinan la simulación computacional intensiva y el análisis de datos masivos).

AMÉRICA LATINA

Brasil

Inria y el LNCC, el Laboratorio Nacional de Computación Científica de Brasil, tienen una larga historia de cooperación científica. En el año 2020 se firmó un acuerdo de colaboración en varios campos de investigación, incluida la IA.

7

Bibliografía y Publicaciones Inria: Cifras

Durante el período 2013-2019, los investigadores de Inria publicaron más de ***450 artículos en revistas de IA y más de 1.800 ponencias para conferencias sobre IA.*** Efectivamente, Inria se encuentra entre las 20 primeras entidades del AI Research Ranking de 2019. La edición 2019 del AI Research Ranking analizó las publicaciones en la conferencia anual de Neural Information Processing Systems (NeurIPS) y la International Conference on Machine Learning (ICML). Utilizando las actas de las conferencias de 2019, se analizaron cada uno de los 2.200 trabajos aceptados, recopilaron la lista de autores y sus organizaciones afiliadas y publicaron el ranking de los principales países y organizaciones. Inria ocupa el puesto 16 en la clasificación general de las organizaciones públicas de investigación. Sólo otras tres entidades públicas europeas aparecen en la lista (Oxford University, la ETH y la EPFL).

8

Bibliografía de lectura complementaria

Esta sección contiene otra bibliografía identificada como relevante para la lectura adicional, agrupadas en categorías. No pretende ser exhaustiva, sino que simplemente ofrece algunas lecturas adicionales a las mencionadas en los capítulos anteriores y a las publicaciones de los equipos de proyecto de Inria.

IA genérica

One Hundred Year Study on Artificial Intelligence (AI100), Stanford University, agosto de 2016, https://ai100.stanford.edu.

AI for humanity. French strategy for AI. https://www.aiforhumanity.fr/en/

Alan Turing. *Intelligent Machinery, a Heretical Theory*. Philosophia Mathematica (1996) 4 (3): 256-260. Artículo original de 1951.

Yves Caseau et al., Renouveau de l'Intelligence artificielle et de l'apprentissage automatique, Commission technologies de l'information et de la communication, Rapport de l'Académie des technologies, 2018

Ernest Davis and Gary Marcus. *Commonsense Reasoning and Commonsense Knowledge in Artificial Intelligence*. Communications of the ACM Vol. 58 No. 9. 2015

Olivier Ezratty, Les usages de l'intelligence artificielle, edición de 2020, que se puede descargar en http://www.oezratty.net/

Michael A. Goodrich and Alan C. Schultz. Human–Robot Interaction: A Survey. Foundations and Trends® in Human–Computer Interaction Vol. 1, No. 3 (2007) 203– 275

Jonathan Grudin. AI and HCI: *Two Fields Divided by a Common Focus*. AI magazine, 30(4), 48-57. 2008

Kevin Kelly. *The Three Breakthroughs That Have Finally Unleashed AI On The World*. http://www.wired.com/2014/10/future-of-artificial-intelligence. 2014

Yang Li, Ranjitha Kumar, Walter S. Lasecki, Otmar Hilliges. Artificial Intelligence for HCI: A Modern Approach. CHI, 2020.

Pierre Marquis, Odile Papine, Henri Prade (eds). *Panorama de l'Intelligence Artificielle. ses bases méthodologiques, ses développements*. 3 vols. Cepaduès. 2014.

Raymond Perrault, Yoav Shoham, Erik Brynjolfsson, Jack Clark, John Etchemendy, Barbara Grosz, Terah Lyons, James Manyika, Saurabh Mishra, and Juan Carlos Niebles, *The AI Index 2019 Annual Report*, AI Index Steering Committee, Human-Centered AI Institute, Stanford University, Stanford, CA, diciembre de 2019.

Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. http://aima.cs.berkeley.edu/

Terry Winograd. *Shifting viewpoints: Artificial intelligence and human–computer interaction*. Artificial Intelligence 170(18):1256-1258. 2006.

Debates sobre la IA

Dario Amodei, Chris Olah et al, Concrete Problems in AI Safety, arXiv:1606.06565v2, 2016

Ronald C. Arkin. *The Case for Ethical Autonomy in Unmanned Systems*. Journal of Military Ethics 12/2010; 9(4)

Anne Bouverot, Thierry Delaporte et al., Algorithmes : contrôle des biais, S.V.P., Institut Montaigne, 2020

Bertrand Braunschweig and Malik Ghallab, editors, Reflections on AI for Humanity, book to be published, Springer, 2020

Erik Brynjolfsson, Daniel Rock and Chad Syverson, Artificial intelligence and the modern productivity paradox: a clash of expectations and statistics Working Paper 24001 http://www.nber.org/papers/w24001

Samuel Butler. *Erewhon*. Free eBooks en Planet eBook.com, 1872.

Lettre du CICDE N°10. Emploi opérationnel de l'intelligence artificielle. April 2018. https://www.irsem.fr/data/files/irsem/documents/document/file/2934/20180412NP-CICDE-Lettre-CICDE-AVRIL-2018.pdf

Kate Crawford, Roel Dobbe, Theodora Dryer et al. AI Now 2019 Report. AINow Institute, 2019, https://ainowinstitute.org/AI_Now_2019_Report.html

Dominique Cardon. *A quoi rêvent les algorithmes*. Seuil, 2015.

Dominique Cardon, Jean-Philippe Cointet and Antoine Mazières, La revanche des neurones, L'invention des machines inductives et la controverse de l'intelligence artificielle, La Découverte «Réseaux» 2018/5 n° 211, pp 173-220, 2018

Thomas G. Dietterich and Eric J. Horvitz. Rise of Concerns about *AI: Reflections and Directions*. Communications of the ACM | Octubre de 2015 Vol. 58 No. 1

Virginia Dignum, Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way, Springer, 2019.

Jessica Fjeld, Nele Achten et al., Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI, https://cyber.harvard.edu/publication/2020/principled-ai, 2020

Carl Benedikt Frey and Michael A. Osborne, The future of employment: how susceptible are jobs to computerisation ?, 2013

Malik Ghallab, Responsible AI: Requirements and Challenges, by request to the author, LAAS-CNRS, University of Toulouse, malik.ghallab@laas.fr, 2020

Thilo Hagendorff. The Ethics of AI Ethics -- An Evaluation of Guidelines. Minds & Machines, 2020.

High Level Expert Group on AI. Ethics guidelines for trustworthy AI. 2019. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthyai

Alexandre Lacoste, Alexandra Luccioni, Victor Schmidt, Thomas Dandres. Quantifying the Carbon Emissions of Machine Learning. 2019 https://arxiv.org/abs/1910.09700

OECD (2019); Deliberations of the Expert Group on Artificial Intelligence at the OECD (AIGO); disponible en https://www.oecd-ilibrary.org/
Stuart Russell. *Human compatible, AI and the problem of control*. Penguin Books, 2019.

Roy Schwartz, Jesse Dodge, Noah A. Smith, Oren Etzioni. Green AI. 2019 https://arxiv.org/abs/1907.10597

Ion Stoica, Dawn Song, Raluca Ada Popa, David A. Patterson, Michael W. Mahoney, Randy H. Katz, Anthony D. Joseph, Michael Jordan, Joseph M.

Hellerstein, Joseph Gonzalez, Ken Goldberg, Ali Ghodsi, David E. Culler and Pieter Abbeel. *A Berkeley View of Systems Challenges for AI*. EECS Department, University of California, Berkeley, 2017.

UNESCO (2019); Preliminary Study on the Ethics of Artificial Intelligence. SHS/COMEST/EXTWG-ETHICS-AI/2019/1; Disponible en https://unesdoc.unesco.org/ Moshe Vardi. On Lethal Autonomous Weapons. Communications of the ACM, diciembre de 2015 vol. 58 no. 12,

Aprendizaje automático

Martin Abadi et al. *Large-Scale Machine Learning on Heterogeneous Distributed Systems*. Software available from tensorflow.org. 2015.

Nicholas Ayache. AI and Healthcare: towards a Digital Twin?. *MCA 2019 - 5th International Symposium on Multidiscplinary Computational Anatomy*, 2019 https://issuu.com/univ-cotedazur/docs/ayache-ai-summit-2018-vl10-uca

Alejandro Barredo Arrieta and Natalia Díaz-Rodríguez and Javier Del Ser and Adrien Bennetot and Siham Tabik and Alberto Barbado and Salvador García and Sergio GilLópez and Daniel Molina and Richard Benjamins and Raja Chatila and Francisco Herrera. *Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI*. Information Fusion, 2020.

Valérie Beaudouin, Isabelle Bloch, David Bounie, Stéphan Clémençon, Florence d'Alché-Buc, et al., Flexible and Context-Specific AI Explainability: A Multidisciplinary Approach, Hal-02506409, 2020

Tarek R. Besold et al., Neural-Symbolic Learning and Reasoning: a Survey and Interpretation, arXiv:1711.03902v1, 2017

Christopher Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

Léon Bottou: *From machine learning to machine reasoning: an essay*, Machine Learning, 94:133-149, enero de 2014.

Mathieu Causse, Cameron James, Mohamed Masmoudi and Houcine Turki,Parsimonious Neural Networks, Adagos Company, 2019

Pedro *Domingos. The Master Algorithm:* How the Quest for the Ultimate Learning Machine Will Remake Our World. Penguin Books, 2015.

Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Franco Turini, Fosca Giannotti, and Dino Pedreschi. *A Survey of Methods for Explaining Black Box Models*. ACM Comput. Surv. 2018.

Leilani H. Gilpin, David Bau, Ben Z. Yuan, Ayesha Bajwa, Michael Specter, Lalana Kagal. *Explaining Explanations: An Overview of Interpretability of Machine Learning*. 2019 https://arxiv.org/abs/1806.00069

Demis Hassabis, Dharshan Kumaran, Christopher Summerfield and Matthew Botvinick, Neuroscience-Inspired Artificial Intelligence, Neuron 95, pp. 245-258, 2017

Michael I. Jordan and Tom M. Mitchell. *Machine learning: Trends, perspectives, and prospects*. Science, Vol 349 Issue 6245. 2015.

Peter Kairouz, H. Brandan MacMahan et al., Advances and Open Problems in Federated Learning, arXiv:1912.04977v1, 2019

Nan Rosemary Ke et al., Learning neural causal models from unknown interventions, arXiv:1910.01075v1, 2019

Yann Le Cun. *The Unreasonable Effectiveness of Deep Learning*. Facebook AI Research & Center for Data Science, NYU. http://yann.lecun.com, 2015

Yann Le Cun. Quand la machine apprend, La révolution des neurones artificiels et de l'apprentissage profond (French). Odile Jacob, 2019.

Volodymyr Mnih et al. *Human-level control through deep reinforcement learning*. Nature 518, 529–533. 2015

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, et al.. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, Microtome Publishing, 2011.

Jonas Peters, Dominik Janzing, and Bernhard Schölkopf, Elements of Causal Inference: Foundations and Learning Algorithms, MIT Press, 2017

David Rolnick et al., Tackling Climate Change with Machine Learning, arXiv:1906.05433v1, 2020

Ribana Roscher, Bastian Bohn, Marco F. Duarte, Jochen Garcke. *Explainable Machine Learning for Scientific Insights and Discoveries*. IEEE Access, 2020.

Bernhard Schölkopf, Causality for machine learning, arXiv:1911.10500v1, 2019

Michèle Sebag. *A tour of Machine Learning: an AI perspective*. AI Communications, IOS Press, 2014, 27 (1), pp.11-23.

Thomas Serre. *Deep Learning: The Good, the Bad, and the Ugly*. Annual Reviews, 2019 Emma Strubell Ananya Ganesh Andrew McCallum, Energy and Policy Considerations for Deep Learning in NLP, arXiv:1906.02243v1, 2019

Deqing Sun, Xiaodong Yang, Ming-Yu Liu y Jan Kautz, PWC-Net: CNNs for Optical Flow Using Pyramid,Warping, and Cost Volume, arXiv:1709.02371v2, 2017

Neil C. Thompson et al, The Computational Limits of Deep Learning, arXiv:2007.05558v1, 2020

Visión

Nicholas Ayache. *Des images médicales au patient numérique*, Leçons inaugurales du Collège de France. Collège de France / Fayard, March 2015.

Yasutaka Furukawa, Jean Ponce. *Accurate, Dense, and Robust Multiview Stereopsis*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010.

Sancho McCann, David G. Lowe. *Efficient Detection for Spatially Local Coding*. Lecture Notes in Computer Science Volume 9008 pp 615-629. 2015.

Farhood Negin, Serhan Cosar, Michal Koperski, François Bremond. *Generating Unsupervised Models for Online Long-Term Daily Living Activity Recognition*. Asian conference on pattern recognition (ACPR 2015), 2015.

A. Rosenfeld, R. Zemel, J.K. Tsotsos, The Elephant in the Room, 2018 https://arxiv.org/abs/1808.03305

Oriol Vinyals, Alexander Toshev, Samy Bengio & Dumitru Erhan. *Show and Tell: A Neural Image Caption Generator*, 2015. https://arxiv.org/pdf/1502.03044. 2015

Representación del conocimiento, web semántica, datos

Bettina Berendt, Fabien Gandon, Susan Halford, Wendy Hall, Jim Hendler, Katharina Kinder-Kurlanda, Eirini Ntoutsi, and Steffen Staab. *Web Futures: Inclusive, Intelligent, Sustainable, The 2020 Manifesto for Web Science*, Dagstuhl Manifesto, pp. 1–44, issn:2193-2433 https://www.webscience.org/wpcontent/uploads/sites/117/2020/07/main.pdf

Tim Berners-Lee, James Hendler and Ora Lassila. *The Semantic Web*. Scientific American, mayo de 2001.

Fabien Gandon. *A Survey of the First 20 Years of Research on Semantic Web and Linked Data*. Revue des Sciences et Technologies de l'Information - Série ISI : Ingénierie des Systèmes d'Information, Lavoisier, 2018.

Fabien Gandon. *The three 'W' of the World Wide Web callfor the three 'M'of a Massively Multidisciplinary Methodology*. Valérie Monfort; Karl-Heinz Krempels. 10th International Conference, WEBIST 2014, Barcelona, Spain. Springer International Publishing, 226, Web Information Systems and Technologies. 2014

Janowicz, K.; Hitzler, P.; Hendler, J.; and van Harmelen, F. *Why the Data Train Needs Semantic Rails.* AI Magazine, 36(5-14). 2015

Antonella Poggi et al. *Linking Data to Ontologies*. Journal on data semantics X Pages 133-173. Springer-Verlag Berlin, Heidelberg. 2008

Robótica y automóviles auto-conducidos

Safety First for Automated Driving – a new cross-industry white paper, 2019. https://www.bmwgroup.com/en/company/bmw-group-news/artikel/Safety-Firstfor-Automated-Driving.html

Jean-François Bonnefon, Iyad Rahwan, and Azim Shariff. *The social dilemma of autonomous vehicles*. Science (2016), 352 (6293). p. 1573-1576.J.

Antoine Cully, Jeff Clune, Danesh Tarapore & Jean-Baptiste Mouret. *Robots that can adapt like animals*. Nature Vol 521 503-507. 2015.

Ethics Commission of the Federal Ministry of Transport and Digital Infrastructure of Germany, Automated and Connected Driving Report, 2017

Christian Gerdes, Sarah M. Thornton. *Implementable Ethics for Autonomous Vehicles*. Autonomes Fahren: Technische, rechtliche und gesellschaftliche Aspekte. Springer, Berlin. 2015.

Pierre-Yves Oudeyer. *Developmental Robotics*. Encyclopaedia of the Sciences of Learning, N.M. Seel ed., Springer References Series, Springer. 2012.

IA y cognición

Stanislas Dehaene, Apprendre !: Les talents du cerveau, le défi des machines (French). Odile Jacob sciences, 2018

Jacqueline Gottlieb, Pierre-Yves Oudeyer, Manuel Lopes and Adrien Baranes. *Information-seeking, curiosity, and attention: computational and neural Mechanisms*. Trends in Cognitive Science (2013) 1-9. 2013.

Douglas Hofstadter & Emmanuel Sander. *L'analogie, cœur de la pensée*. Ed. Odile Jacob, 2013.

Daniel Kahneman. *Thinking, Fast And Slow. New York*: Farrar, Straus and Giroux, 2011

Luc Steels. *Self-organization and selection in cultural language evolution*. In Luc Steels (Ed.), Experiments in Cultural Language Evolution, 1 – 37. Amsterdam: John Benjamins. 2012.

Lenguaje natural, habla, audio

Daniel Adiwardana et al., Towards a Human-like Open-Domain Chatbot, arXiv:2001.09977v1, 2020

Louis Martin, Benjamin Muller, Pedro Javier Ortiz Suárez, Yoann Dupont, Laurent Romary, et al.. CamemBERT: a Tasty French Language Model. 2019.

Kenneth Church. *A Pendulum Swung Too Far*. Linguistic Issues in Language Technology – LiLT. Volume 2, Issue 4. 2007

G. Hinton, L. Deng, D. Yu, G.E. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T.N. Sainath, B. Kingsbury, *Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups*. IEEE Signal Processing Magazine, 29(6):82-97, 2012.

Alec Radford, Karthik Narasimhan, Tim Salimans and Ilya Sutskever. *Improving Language Understanding by Generative Pre-Training*. OpenAI, 2018. https://s3-uswest-2.amazonaws.com/openai-assets/research-covers/languageunsupervised/language_understanding_paper.pdf

Stephen Roller et al., Recipes for building an open-domain chatbot, arXiv:2004.13637v2, 2020

Ashish Vaswani et al., Attention Is All You Need, 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, arXiv:1706.03762v5, 2017

Domaine de Voluceau, Rocquencourt BP 105
78153 Le Chesnay Cedex, France
Tél. : +33 (0)1 39 63 55 11
www.inria.fr