

PABLO RODRÍGUEZ

PRÓLOGO DE SANDY PENTLAND, PROFESOR DEL MIT MEDIA LAB



INTELIGENCIA ARTIFICIAL

CÓMO CAMBIARÁ EL MUNDO
(Y TU VIDA)



En la actualidad, la recolección y análisis de datos ha adquirido un papel esencial en nuestras vidas, no sólo para otorgar una imagen certera de una situación concreta sino también para predecir actuaciones o posibilidades futuras. Los datos y la inteligencia artificial existen prácticamente en todo lo que usamos y hacemos. Están en teléfonos, coches, carreteras, líneas eléctricas, cursos de agua, contenedores de comida y una cantidad innumerable de otros elementos. Además, sus posibilidades son inmensas, los macrodatos y la IA pueden ayudarnos a predecir variables como los precios de mercado, las sequías, las migraciones o una futura escasez de alimentos.

En este libro, Pablo Rodríguez realiza un análisis de las dimensiones que los macrodatos y la IA están adquiriendo en nuestra sociedad y como están transformado aspectos como la seguridad, la salud, las finanzas o el propio aprendizaje.



Pablo Rodríguez

Inteligencia artificial

Cómo cambiará el mundo (y tu vida)

ePub r1.2

Titivillus 29.03.2020

Título original: *Inteligencia artificial*
Pablo Rodríguez, 2018
Traducción: Jorge Rizzo

Editor digital: Titivillus
ePub base r2.1



Índice de contenido

Inteligencia artificial

Prólogo

Introducción

1 El regreso al futuro de los macrodatos

¿Por qué deberías preocuparte?

El futuro empieza ahora

Un viaje al interior de los datos

Historia antigua de los datos

La emergencia de la estadística

Los primeros días del almacenaje moderno de datos

Los inicios de la inteligencia empresarial

La aparición de los grandes centros de datos

La aparición de internet

Las primeras ideas sobre macrodatos

La web 2.0 aumenta los volúmenes de datos

Se adopta el uso actual del término «big data» (macrodatos)

El poder de las redes de datos

Interconectando Datos: propiedades mágicas

Científicos de datos: los gurús de los macrodatos

2 El poder de las cosas y sus datos

La Internet de las Cosas

Cómo acabar con internet

Una distribución de datos más rápida

La IdC a la velocidad de la luz

La Internet de las Cosas y los datos

Seguridad para todos

3 Inteligencia artificial: humanos contra máquinas

IA

Algoritmos

Aprendizaje automático

Aprendizaje cognitivo

La IA te cambia la vida

¿Te da miedo la IA?

La IA y los macrodatos

4 La IA para el liderazgo y la creatividad

Liderazgo y fútbol: el colectivismo y el deporte

El fútbol como algoritmo de datos
Un laboratorio de datos en el campo
El ordenador en la cocina
El lenguaje de los sabores
La IA y la gastronomía

5 Apostarlo todo a la ciencia de datos y a la IA
Un paseo por La Boqueria
La IA y el turismo
La IA y la seguridad
La IA y las finanzas
La IA, el empleo y la movilidad

6 La IA en busca del bien global
Un día en Harvard
La IA y la salud mental
La salud física y la IA
La IA para el bien global

7 IA ética, privacidad y el alma de los datos
La utopía del ciberespacio
La privacidad de los datos
Un delicado equilibrio
Bancos de datos personales

Conclusión

Bibliografía

Agradecimientos

Autor

Para Uriel

Prólogo

Casi todos los problemas importantes del ser humano se pueden abordar usando datos y la inteligencia artificial (IA). Pero históricamente ha sido imposible, dada la falta de referencias y la falta de enlaces que permitan el análisis de los datos.

En la última década se ha registrado un cambio enorme en el mundo de la IA y los datos. Quizá el mayor avance de todos sea que ahora el 80 por ciento de la totalidad de los seres humanos tenemos acceso a un teléfono y todo el mundo está conectado digitalmente al resto del mundo. Hasta las comunidades más pobres tienen acceso a las telecomunicaciones. Y los registros de esas interacciones miden aspectos importantes de la sociedad humana que los utiliza, porque los datos que generan al conectarse entre ellos o con las redes de telecomunicaciones, nos aportan cifras de población, patrones de movimiento y patrones de interacción entre comunidades. Y como consecuencia, ahora disponemos de informaciones sobre las condiciones de vida de los seres humanos en casi todos los rincones del mundo, y sobre los enlaces de conexión usados para obtener esos datos.

Esta nueva infraestructura de comunicaciones nos ha aportado posibilidades absolutamente impresionantes para utilizar la IA. Los registros nos dan todo tipo de información sobre nuestro entorno inmediato y podemos interactuar con ellos con la IA de un modo que hace diez años habríamos considerado una locura o algo carísimo. ¿Quieres saber cómo está el tráfico de camino a casa? ¿Qué restaurantes les gustan a tus amigos? ¿Qué tiendas del barrio están abiertas? Sólo hay que consultar el *smartphone*. Es algo asombroso.

Las últimas dos décadas, mi equipo del MIT Media Lab y yo hemos estudiado cómo nos afecta esto a nosotros y a nuestras interacciones, y cómo influyen estos efectos en los grandes cambios registrados por la sociedad. Por ejemplo, a través de la IA examinamos el patrón de interacción en comunidades, y nos preguntamos si los miembros de las comunidades hablan entre ellos, si hablan o no con el resto de la sociedad, si exploran o hablan con personas nuevas, o si hablan una y otra vez con las mismas personas.

Estudiando estos patrones hemos observado algo realmente sorprendente: se puede comprender la esencia de una comunidad, esas características humanas que antes eran imposibles de cuantificar, simplemente midiendo sus patrones de interacción y la IA. Los patrones de interacción nos cuentan si la comunidad está sana o enferma, si es pobre o rica, si es innovadora o si no avanza. Todo eso puede verse en el patrón de interacción entre las personas y con algoritmos de IA. Porque no es la tecnología la que dicta la esencia de la comunidad; son las personas y son las ideas que fluyen entre las personas lo que diferencia a una comunidad de otra.

Por otra parte, dándole la vuelta a este hallazgo científico podemos analizar el patrón de interacciones de esa comunidad a través de los coches o las bicicletas que circulan, a través del uso de las tarjetas de crédito o de los registros gubernamentales, y a partir de esos patrones podemos medir el lado humano de la comunidad. A partir de esas miguitas digitales y con IA se pueden identificar bolsas de pobreza, se pueden identificar barrios con un mayor riesgo de delincuencia, o barrios que se están convirtiendo en centros de innovación.

Por ejemplo, en Costa de Marfil, donde han sufrido recientemente una guerra civil, y donde no tienen censos, hemos visto que usando referencias de patrones de comunicación e IA se puede determinar qué poblaciones son pobres y cuáles son ricas. Y hacer este tipo de análisis sale barato. Eso significa tener un censo de la pobreza a tiempo real por sólo unos céntimos. Del mismo modo, se puede ver qué barrios tienen más posibilidades de convertirse en fuentes de innovación porque comunican mejor, y cuáles están en riesgo de registrar un aumento de la delincuencia porque cada vez están más aislados de las comunidades de los alrededores.

En realidad impresiona ver a través de la IA y la ciencia de los datos, que los patrones de comunicación son, aparentemente, el factor más importante a la hora de determinar las características humanas de una comunidad. Según parece no es el nivel de educación lo que más importa, ni tampoco el marco regulatorio: son las interacciones humanas. Antes de disponer de este tipo de referencias no entendíamos la importancia de los patrones de interacción humana. Había muchas teorías sobre por qué es diferente la delincuencia en distintos lugares, pero no tuvimos la respuesta hasta que empezamos a disponer de esas miguitas digitales que nos han permitido analizar los patrones reales de interacción.

Este cambio de «La mitad del mundo no ha hecho nunca una llamada telefónica» a «Podemos disponer de mapas de pobreza y de delincuencia a tiempo real» es lo que el secretario general de la ONU llama la Revolución de los Datos y la IA. Por primera vez en la historia del mundo tenemos realmente la posibilidad de ver lo que sucede en el mundo en forma de datos.

Durante toda la historia de la humanidad hemos sufrido epidemias que han diezmando la población, y los gobernantes, desde su capital, no se enteraban siquiera de lo ocurrido hasta después de que hubiera muerto la gente. En 1918, por ejemplo, hasta las sociedades más avanzadas hacían el seguimiento de la mortífera gripe registrando el número de cuerpos que llegaban al depósito. Pero ahora, gracias a una tecnología digital móvil que está por todas partes, a unos potentes ordenadores y a unas elaboradas técnicas estadísticas, podemos detectar las primeras señales de estos desastres e irlos siguiendo a medida que se desarrollan. Por ejemplo, la próxima vez que haya una epidemia de ébola, podremos usar montones de miguitas digitales y IA para hacer el seguimiento, frenarla y contenerla.

La Asamblea General de las Naciones Unidas ha decidido que los Institutos Nacionales de Estadística de cada país usen este tipo de métodos de análisis de

informaciones digitales de forma continuada por el bien de la sociedad. Eso significa que los 193 países miembros de la Asamblea General de la ONU deben comprometerse a que sus Institutos de Estadística comuniquen estas mediciones de «datos para el bien social» de modo que el mundo pueda tener más información sobre violencia, desigualdades, calidad de vida, propagación de enfermedades, etc., de forma más o menos continuada. Y así, por primera vez, podemos imaginarnos un mundo en el que haya de verdad transparencia y en que se responda del funcionamiento de las políticas de los gobiernos y de las intervenciones sociales, así como del alcance de la ayuda al desarrollo.

Estamos tan acostumbrados a que la situación real quede oculta tras una nube opaca que nos cuesta imaginar un mundo en el que por todas partes se hagan mediciones fiables y a tiempo real sobre la condición humana. Pero lo cierto es que estamos progresando en dirección a ese objetivo. Por supuesto, todos estos datos que podrían proporcionarnos recursos prácticos pero también transparencia, responsabilidad, un mejor gobierno y un mundo que tenga en cuenta a todos y cada uno de los seres humanos y se preocupe por ellos, también tienen un aspecto negativo.

El aspecto negativo son los delincuentes informáticos y las malas praxis comerciales, que no respetan las normas básicas de privacidad de datos, y a muchos les preocupa este lado oscuro del uso de los registros. Telefónica, por ejemplo, ha creado un índice con el que pregunta a personas de todo el mundo qué les parecen sus registros, y han observado que un 25 por ciento más o menos está de acuerdo con la ecología de datos actual. No obstante, otro 25 por ciento manifiesta serias dudas de que vayan a producir resultados positivos, y cree que sus informaciones, sumados al impacto de los algoritmos de la inteligencia artificial que los procesan, pueden provocar resultados negativos.

Así pues, ¿qué vamos a hacer con todos esos datos? Bueno, yo creo que tenemos que promover el uso de éstos con fines positivos, para comprendernos mejor a nosotros mismos, comprender nuestras comunidades y ayudar al mundo en general. Pero también tenemos que ser prudentes; tenemos que hacer los deberes. Pablo Rodríguez, autor de este libro, fue fundamental para que mi equipo del MIT, la Mozilla Foundation, el Open Data Institute del Reino Unido y Telefónica combinaran esfuerzos en la creación del Data Transparency Lab, para que los investigadores de todo el mundo puedan comprender mejor lo que se hace con nuestros registros, los algoritmos de IA y cómo hacer el mejor uso posible de ellos.

Este libro promueve esta actitud, el uso de los datos y la IA para el bien personal y para el bien de la sociedad, pero comprendiendo al mismo tiempo los peligros de un mundo rico en éstos, y cómo se puede construir un mundo que nos proteja de estos peligros. Los datos tienen poder. Los datos son valiosos. Este libro explora el potencial de éstos y de la inteligencia artificial, sus aplicaciones para el individuo y su potencial para transformar el mundo por completo, con una llamada a la prudencia

para evitar las trampas que se nos presentan, de modo que todos podamos disfrutar de un planeta mejor, con menos desigualdades y más oportunidades para todos.

PROF. SANDY PENTLAND
MIT Media Lab

Introducción

Corría el año 1997, en un laboratorio en el sótano del King's College de Londres. En el momento de entrar, estaba lloviendo y, siendo Londres, probablemente seguiría lloviendo en aquel momento. También habría oscurecido. Pero no había ventanas, así que no podía ver el exterior, y no había hablado con nadie desde mi llegada, salvo con el profesor Alan Rogers, supervisor de mi tesis. Llevaba trabajando sin hacer una pausa desde el momento en que había dejado mis cosas en la taquilla y cerrado la puerta del laboratorio. Estaba trabajando en un experimento relacionado con mi tesis, mediciones de alta resolución usando técnicas fotónicas. Parte de mi trabajo consistía en manipular lentes y láseres pero, sobre todo, implicaba estar sentado frente al terminal del ordenador jugando con algoritmos para contar fotones que volaban por ahí para poder señalar cambios de temperatura muy precisos que nos permitieran detectar tumores en fases muy tempranas.

Pero cada vez le encontraba menos encanto al trabajo de física y fotónica. En lugar de usar mi ordenador para procesar números, empecé a hacer pruebas con él, simplemente para ver cómo funcionaba y experimentar. Este ordenador había cambiado radicalmente mi modo de trabajar. Cuando estudiaba en España usábamos un antiguo lenguaje de programación llamado Fortran. Los datos se almacenaban en discos flexibles del aspecto de posavasos en los que sólo cabían migajas de ellos en comparación con lo que acumula hoy en día un reloj digital. Muchas veces los dejábamos trabajando toda la noche para que completaran los cálculos, y volvíamos al día siguiente como niños esperando los regalos del *tió* (El *tió* es una peculiar tradición catalana consistente en un tronco que es «alimentado» con golosinas por los niños durante el adviento. ¡En Navidad le dan golpes con palos mientras cantan una canción y el tronco «caga» regalos!).

En términos tecnológicos, los ordenadores que usé en España durante la carrera universitaria eran tan avanzados como troncos de madera en comparación con el ordenador con el que estaba jugando ahora en Londres. Y cada año los ordenadores se volvían más pequeños, más baratos, más potentes y más rápidos.

Yo también había ido cambiando a un ritmo similar. Hasta aquel momento me contentaba con trabajar solo en mis experimentos, dejando volar la imaginación a través de los algoritmos, y trasteando entre lentes y láseres con los dedos. Durante mi juventud en mi Asturias natal me sentía un bicho raro, un introvertido en tierra de extrovertidos. Los españoles, y los asturianos en particular, suelen ser tipos gregarios, que buscan el ruido y la compañía. Si entráis en cualquier bar de Oviedo os encontraréis gente charlando, haciendo esfuerzos por hacerse oír con el ruido del televisor de la esquina y la radio a todo volumen.

No es que fuera tímido. Es que me encantaba mi trabajo. Siempre me ha gustado disfrutar del placer de descubrir cosas nuevas. Pero el tipo de trabajo que estaba haciendo en Londres, pasando horas y horas solo haciendo cálculos matemáticos y experimentos, ya no me provocaba ninguna emoción. Había otra cosa que me llamaba más la atención: el ordenador e internet. Esa poderosa herramienta que podía hacer tantos cálculos, permitiéndome dejar volar la imaginación, en lugar de empantanarme en cálculos que ahora una máquina podía hacer por mí. Aquello suponía experimentar los efectos de tener una inteligencia artificial y aumentada a mi disposición.

Con el descubrimiento de los ordenadores y de internet cambió algo más. Descubrí gente muy interesantes. Extraño, ¿no? Normalmente tenemos esa imagen de los ingenieros informáticos trabajando solos por la noche, con el pálido rostro iluminado por el brillo de la pantalla. En aquella época, la de los albores de internet, no era así en absoluto. Era una comunidad de personas, en su mayoría científicos e ingenieros como yo (que ahora se llaman científicos informáticos), todos trabajando en una cosa completamente nueva, algo cuyo significado quizá sólo unos pocos de ellos entendían realmente. El objeto de su estudio era internet y la web, una red de comunicaciones, un almacén de conocimiento que acabaría volviéndose omnipresente en el mundo en una década.

Aquello creció tan rápidamente que, igual que un joven que de pronto crece quince centímetros, el estirón empezó a provocarle dolores. Estábamos estirando internet y llevándola al límite, y eso es en lo que yo trabajaba: en hacer que internet fuera más rápida y más eficiente para la cantidad de registros que transportaba, que aumentaba exponencialmente (desde complejas páginas web hasta vídeos en directo, o parches antivirus para evitar ciberataques).

Trabajé en algunos de los lugares más innovadores y pioneros de la investigación de internet, en Microsoft, en los Bell Laboratories y en el Instituto Federal Suizo de Tecnología. Descubrí y acabé formando parte de una comunidad de investigadores, científicos e ingenieros de datos y algoritmos que iban prácticamente improvisando, diseñando, construyendo y manteniendo algo nunca visto. Una comunidad que formaba parte de la historia de internet, que la transformó de un servicio lento y experimental a una maravilla a la velocidad de la luz. Fue una época emocionante; una época de muchísimo trabajo, de horarios infinitos, de grandes amistades y de innovaciones vertiginosas.

A los pocos años, ya hacía tiempo que había dejado el trabajo solitario del laboratorio londinense en el King's College, trabajando sobre oscuros problemas de la física teórica. Estaba creando y construyendo nuevos negocios. Creábamos y vendíamos compañías tecnológicas en Silicon Valley. Era profesor asociado en grandes universidades estadounidenses. Y aún no había cumplido los cuarenta años. Pero se acercaba otra transformación (y sin duda habrá más en el futuro). Y esa transformación es el objeto de este libro. Fue entonces cuando volví a España (¡y al

sol!) y (re)conecté con una comunidad, una comunidad globalizada de líderes en campos variados como la medicina, el liderazgo, la creatividad o la expendeduría.

Siempre he querido resolver problemas. Primero fueron teóricos; luego fueron tecnológicos. Ahora, con todo lo que he aprendido y con un extraordinario equipo, me dedico a problemas que pueden tener un tremendo impacto en la sociedad y que requieren grandes avances tecnológicos. ¿Qué pueden aportar los ingenieros informáticos como yo a problemas como la depresión? ¿O la pandemia de crisis global? ¿O la delincuencia? ¿O el liderazgo? ¿O el turismo?

He pasado gran parte de mi vida profesional estudiando y diseñando soluciones para enormes redes de telecomunicaciones. Redes que se extienden por continentes enteros y que sólo en los últimos cinco años han transportado más datos que todos los generados por la humanidad desde el inicio de la historia. Pero me he dado cuenta, a través de conversaciones, reuniones casuales y a veces por simples ideas que se me han ocurrido, que hoy en día el conocimiento y la experiencia que han ganado los ingenieros informáticos durante los vertiginosos años de aparición de internet y con los recientes avances en *big data* e inteligencia artificial (IA) pueden aplicarse a otras disciplinas y al nivel personal, con resultados extraordinarios. Es como enrollar a un constructor de barcos en un proyecto de diseño de una catedral (algo que en realidad ya se está haciendo, en los trabajos de construcción de la Sagrada Familia de Barcelona). No solo supone la incorporación de otro punto de vista, sino también de herramientas específicas y procesos desarrollados para un campo pero adaptables a otro.

Y esta influencia de diferentes tecnologías, los datos y la inteligencia artificial en la sociedad están creando grandes cambios. La mayoría del dinero que has gastado en los últimos veinticinco años en libros, ocio, educación, información salud y transporte se está desmaterializando gracias a la tecnología, los ordenadores e internet, que hacen que tu dinero llegue más lejos para que puedas usarlo con otros fines. La tecnología está haciendo posible que todas estas necesidades cuesten mucho menos, probablemente lo que cuesta una conexión a internet. Muy pronto contarás con los mejores laboratorios de análisis médicos en el teléfono, a través de sensores que detectarán y analizarán tus signos vitales o te analizarán el ADN; no será necesario trasladarse tanto para acceder a los mejores médicos y especialistas; ya tienes todas las enciclopedias posibles en la punta de los dedos, acceso a las noticias, información y ocio, y los sistemas de computación más potentes a tu alcance. Recuperarás tiempo libre para hacer mejor lo que más te guste y lo que mejor se te dé, ganando así tiempo para hacer más actividades cada vez y mejorar exponencialmente.

Internet ha hecho que dispongamos de datos en abundancia, que estén por todas partes y que sean mucho más valiosos. Tanto si sales a correr como si ves la tele o simplemente estás sentado en el coche en un atasco, prácticamente cada actividad que desarrollas crea un rastro digital, más materia prima para las destilerías de registros.

Ahora que dispositivos tan diversos como relojes o coches conectan con internet, el volumen va en aumento: hay quien calcula que un coche que se conduzca solo generará 100 gigabytes por segundo.

Por su parte, las técnicas de inteligencia artificial sacan un valor añadido a estas informaciones. Los algoritmos pueden predecir cuándo un cliente está listo para comprar, cuándo hay que revisar un motor a reacción o cuándo una persona corre peligro de desarrollar una enfermedad. Hoy en día, gigantes industriales como General Electric o Siemens se venden como empresas de datos.

Y esto es importante porque tus datos personales contienen tu alma. Éstos son un espejo de ti mismo, de tus emociones. La mayoría piensa que los datos son fríos, impersonales, pero pueden revelar tu lado emocional, pueden ayudarte a crecer. Los datos y la inteligencia artificial se están convirtiendo en el nuevo psicólogo, en el nuevo *coach*; ellos nos ayudan a prevenir nuestras enfermedades, a reaccionar durante desastres naturales y a resolver los problemas sociales más acuciantes.

Este libro cuenta la historia de las personas que he conocido y de las colaboraciones en que he participado para resolver problemas de muy diversos campos, utilizando registros y la inteligencia artificial. Una nueva innovación que crea impacto a nivel global y es sostenible y se preocupa de la conciencia humana. Ahora aquel sótano de Londres queda muy lejos. Visito hospitales, organizaciones internacionales de salud, grandes equipos de liderazgo y creatividad, viajo a las profundidades de los centros de datos y al centro de la red y voy aprendiendo sobre salud, alimentación, energía, sobre economía mundial o sobre lo que hace feliz a la gente. Aplico todo lo que he aprendido, y sigo aprendiendo, para ayudar a resolver los problemas de médicos y enfermos, ingenieros de redes, activistas en defensa de los derechos humanos, organizaciones humanitarias, entrenadores de fútbol, artistas y chefs de cocina, empresas sociales, organizaciones sin ánimo de lucro, o tus propios problemas, para hacer del mundo un lugar mejor. Y todo ello está sucediendo con la revolución del «alma de tus datos», y de la inteligencia artificial.

DOCTOR PABLO RODRÍGUEZ

El regreso al futuro de los macrodatos

Cualquier loco inteligente puede hacer las cosas más grandes, más complejas y más violentas. Pero hace falta un toque de genialidad —y mucho valor— para moverse en la dirección opuesta.

E. F. SCHUMACHER

¿Por qué deberías preocuparte?

Durante la parte más productiva de mi trayectoria como científico, pasé unos años fantásticos en Cambridge, la famosa ciudad universitaria junto al río Cam de la que han salido mentes brillantes como Stephen Hawking o Isaac Newton. Cambridge desprende un aire de placidez y la sensación de que hay espacio para que las cosas maduren en su momento, algo que no he encontrado en ningún otro lugar. Es un lugar que ofrece a la mente el espacio necesario para el pensamiento profundo necesario para poder idear las grandes revoluciones tecnológicas. Y fue en Cambridge donde creció y estudió Alan Turing, el científico informático más decisivo de todos los tiempos. También es el lugar donde vive mi amiga la doctora Anastasia Christofilopoulou. Anastasia lleva cuatro años preparando una exposición sobre criptografía y el Proyecto Enigma, que ayudó a poner fin a la segunda guerra mundial y marcó el inicio de una nueva era de datos, algoritmos e inteligencia computacional nunca vista antes.

Vive con su marido Christos Gkantsidis en Ely, donde me invitan a algo tan inglés como el té en una tarde de lluvia, en pleno otoño, para preparar nuestra reunión familiar anual en algún rincón soleado del Mediterráneo. Ely es un bonito pueblo al norte del condado de Cambridgeshire, con un fantástico paseo junto al río y un animado centro donde se celebra cada año el Festival de la Manzana para la promoción de este fantástico producto nacional. La exposición en la que está trabajando Anastasia es una muestra pionera e interdisciplinaria. *Codebreakers and Groundbreakers* unirá, por primera vez, los notables logros intelectuales y las narrativas paralelas de dos grupos de criptógrafos (*codebreakers*) que trabajaban a la vez, pero de forma independiente: los que intentaban descifrar los códigos de la segunda guerra mundial y los que descifraron el Lineal B, el primer sistema de escritura comprensible de Europa. En una exposición que incluye una máquina Enigma prestada, excepcionalmente, por el GCHQ (el Cuartel General de Comunicaciones y la Agencia de Inteligencia de El Reino Unido), y documentos únicos archivados en la Universidad de Cambridge. La exposición explica la vida y

conmemora los logros de criptógrafos de la segunda guerra mundial como Alan Turing o Bill Tutte, así como de Michael Ventris y John Chadwick (también criptógrafo de la base de Bletchley), que descifraron el Lineal B. Es una muestra fascinante, pero yo lo que busco es la historia que desencadenaron los datos, los algoritmos y los descubrimientos en computación que dieron paso a la inteligencia de las máquinas.

Todo empezó en 1951, en Manchester, cuando el M16, servicio de inteligencia británico, interceptó un mensaje que le habían robado a Alan Turing. Alan, conocido profesor de Cambridge, recibió la visita de la policía, que le preguntó por el robo. Le encontraron en casa, pero él mostró una actitud desdeñosa. Los agentes declararon que era una persona insoportable y sospecharon que ocultaba algo.

Años atrás, en 1939, al declararse la guerra, 800.000 personas habían sido evacuadas de Londres. La segunda guerra mundial se había alargado seis años. Las fuerzas aliadas eran limitadas e iban perdiendo efectivos. Estados Unidos y Canadá intentaban colaborar enviando convoyes de material estratégico.

Alan Turing, que entonces tenía veintisiete años, había llegado a Bletchley Park en el tren, escoltado por agentes de la Marina. Esperaba en el despacho del comandante Denniston. Cuando llegó el comandante, Alan tenía frío y no parecía estar de buen humor. El comandante le preguntó por qué quería trabajar para el gobierno; él respondió que no quería. Mencionó que no le interesaba mucho la política, y el comandante dijo que posiblemente aquella sería la entrevista de trabajo más corta de la historia. Alan mencionó que no hablaba alemán, pero le dijo al comandante que era uno de los mejores matemáticos del mundo. Se planteaba los códigos alemanes como rompecabezas, que le gustaba solucionar. El comandante le pidió a su secretario que acompañara a Alan a la salida, a lo que Alan respondió con una palabra: «Enigma», dejando claro que estaba al corriente del programa de alto secreto para el que había sido escogido como candidato. Alan explicó que Enigma era el mecanismo de encriptación más grande de la historia, y que si los Aliados conseguían descifrar el código, aquello pondría fin a la guerra. El comandante le dijo que todo el mundo estaba convencido de que Enigma era indescifrable. Alan le pidió que le dejara intentarlo, para que pudieran estar seguros.

Así pues, le permitieron participar en Enigma junto a otros, como Peter Hilton, John Cairncross o Hugh Alexander. Habían conseguido hacerse con una máquina Enigma auténtica que lograron sacar de Berlín, pero no conocían la configuración necesaria para que descodificara los mensajes. Cada noche, a las doce, los alemanes cambiaban la configuración; como interceptaban el primer mensaje cada mañana a las 6.00, los criptógrafos sólo tenían dieciocho horas al día para intentar descifrar el código antes de que volviera a cambiar, cuando tendrían que volver a empezar de cero. Hugh (Matthew Goode), campeón de ajedrez, calculó que eso significaba que

había 159 trillones de posibilidades al día. Alan no tenía muy claro que debieran trabajar en equipo; Stewart Menzies, el jefe del MI6, les dijo que en los últimos minutos habían muerto cuatro hombres por no haber podido descifrar el código, y les ordenó que se pusieran manos a la obra.

Alan le dijo al equipo que todos los mensajes estaban flotando por el aire, al alcance de cualquiera que quisiera hacerse con ellos. El problema era que estaban encriptados y que había 159 trillones de posibilidades. Llevaría 20 millones de años probarlo todo. Una misión imposible. Pero Alan Turing tenía una idea brillante en mente. Estaba convencido de que reuniendo suficientes referencias de mensajes anteriores, descifrando unas cuantas claves sobre cómo se habían encriptado aquellos mensajes y usando sofisticados algoritmos de predicción irían mucho más rápido y podrían descodificar los mensajes cifrados. Si capturaban una gran cantidad de datos y llegaban a comprender cómo se encriptaban los mensajes, podrían intentar predecir lo que escribían los alemanes. ¡Sí, era el inicio de la era de los macrodatos y la inteligencia artificial!

La información ha pasado de ser escasa a superabundante. Eso genera unas enormes ventajas que influyen en nuestro día a día. Los macrodatos y la IA nos ayudan a ahorrar dinero en lo que comemos con programas de fidelidad, páginas web de reintegros y cupones de descuento, todo ello diseñado para reducir la factura semanal de la compra. Lo mismo se puede aplicar al transporte, al recreo, al ocio y a las vacaciones.

Nuestra forma de comprar *online* ha cambiado para mejor gracias a la emergencia de nuevos algoritmos de proceso de datos. La mayoría de cosas que buscamos en internet o en televisión vienen acompañadas de opiniones de cliente y de análisis más detallados de empresas de valoración de productos. Éstos nos dan una imagen mucho más clara de los productos que compramos y de los que han comprado otras personas como nosotros, y sirven para que nos puedan recomendar nuevos productos.

No son sólo las compras en internet las que se han visto modificadas con los macrodatos y la IA; es toda la red. Descubrir nueva información a través de internet nunca ha sido tan fácil. La población global de internet creció un 14,3 por ciento entre 2011 y 2013, y ahora son 3.000 millones de personas los que tienen acceso a la red.

Las redes sociales son un elemento de los macrodatos que ha observado un rápido crecimiento en los últimos años. La constante emergencia de información actualizada a través de Twitter e Instagram da una accesibilidad increíble a noticias y registros de todo el mundo. Cada minuto se envían 204 millones de emails en todo el mundo, y se cuelgan 277.000 tweets y 216.000 posts de Instagram.

Ninguno de nosotros disfruta recibiendo facturas de suministros, pero los macrodatos pueden ayudar a reducir esos costes. Es mucho más fácil rastrear y

monitorizar el uso que se hace de la energía para poder configurar temporizadores de calefacción o una presión del agua que nos permita ahorrar en la factura. Los macrodatos y la IA también han simplificado la monitorización del estado del coche en cuanto a kilometraje y consumo de gasolina. Muchos coches cuentan ya con tecnología integrada, como sistemas de navegación por satélite o sensores para el aparcado. Si usas el transporte público para ir a trabajar, dispones de informaciones en tiempo real que te permiten planificar tus viajes por si hay retrasos en el servicio o carreteras o líneas de ferrocarril cortadas por obras.

Las autoridades y los servicios de emergencia usan los macrodatos para poder analizar mejor la delincuencia y los accidentes en todo momento, y la recopilación de datos es cada vez más efectiva. Sin duda, es algo que va a imponerse.

Todo lo que hacemos implica un uso de macrodatos e IA. Desde el momento en que enciendes la tele por la mañana hasta que te acuestas por la noche. Y es una tendencia que va acelerándose cada vez más. En los últimos veinte años, las tecnologías inalámbricas y la conexión a internet se han convertido en algo generalizado, asequible y disponible prácticamente para cualquiera. El uso de teléfonos móviles va aumentando exponencialmente, del 2 por ciento en el año 2000 al 28 por ciento en 2009 y el 70 por ciento en 2017. Hoy en día, personas sin educación formal y con poco que comer se conectan por telefonía móvil de un modo impensable hace sólo treinta años.

Años atrás, cuando la revolución de los *gadgets* tecnológicos empezó a arraigar, la gente solía comprarse dispositivos que mejoraran su calidad de vida —como equipos de música, cámaras fotográficas, sistemas de ocio, enciclopedias, etc.—. Comprar todas esas cosas por separado podía costar miles de dólares. Hoy en día, todo eso viene de serie en un *smartphone* que sale por una fracción de ese precio, y todo lo que no esté incluido en el precio se puede comprar en forma de *app* por menos de lo que cuesta un café.

De este modo, la tecnología se introduce rápidamente en nuestras vidas, cambiándonos para siempre en ciclos cada vez más cortos. Hoy en día tenemos que observar lo que puede suceder en los próximos quince años para comprender lo que ocurrirá dentro de cinco. Ahora mismo, una persona en Papúa Nueva Guinea tiene más posibilidades de comunicación telefónica de las que tenía el presidente de Estados Unidos hace veinticinco años. Y si tiene un *smartphone* con acceso a internet, goza de un mejor acceso a la información del que tenía el presidente hace sólo quince años. En otras palabras, vivimos en un mundo de rápida aceleración de la tecnología y los macrodatos.

En los próximos años presenciaremos el desarrollo de nuevas tecnologías transformadoras —sistemas de computación, computación cuántica, realidad virtual, redes de siguiente generación, inteligencia artificial, redes neuronales, robótica, interfaces de usuario, comprensión de la mente, salud predictiva y ordenadores que igualan la capacidad bruta de procesamiento del cerebro—, lo cual muy pronto hará

posible que la gran mayoría de la humanidad experimente muchas cosas que hoy aún están limitadas al reducido grupo de los que tienen una capacidad inmensa de procesamiento de datos.

Lo que significa todo esto es que, por impresionante que resulte la velocidad del cambio tecnológico mundial registrado hasta el momento, aún no hemos visto nada. Cuando todas estas tecnologías (que siguen avanzando a un ritmo exponencial) se combinen, se harán progresos mayores que nunca. El índice actual del progreso tecnológico es más que suficiente para resolver numerosos retos sociales para todos. Esto, combinado con los macrodatos y la IA, significa que muchas preguntas que hasta ahora permanecían sin responder por fin tendrán respuesta.

Los datos y la IA existen prácticamente en todo lo que usamos y hacemos. Están en teléfonos, coches, carreteras, líneas eléctricas, cursos de agua, contenedores de comida y una cantidad innumerable de otros elementos que nunca habríamos identificado con los ordenadores.

Éstos dicen mucho sobre nuestra conducta colectiva y nuestra sociedad, lo que nos permite hacer cosas increíbles. Por ejemplo, podremos usar los macrodatos y la IA para predecir con suficiente antelación la escasez de alimentos. Variables como los precios de mercado, las sequías, las migraciones, la producción regional anterior o las variaciones estacionales influyen en esta clasificación y en el modelo de aprendizaje de la estructura causal para predecir la probabilidad de que un habitante del entorno rural tenga dificultades para obtener alimentos. Podremos usar los macrodatos y la IA para comprender las trampas de pobreza y cómo salir de ellas, o cuantificar una ola de delincuencia antes de que se extienda en determinando el lugar, el momento y la naturaleza del delito, o decidir qué intervención o serie de intervenciones es más conveniente para mejorar el aprendizaje en las escuelas de países en desarrollo.

Ahora que tenemos todo esto al alcance de la mano, hay una sencilla comprobación que suelo usar: ¿Estás trabajando en algo que pueda mejorar el mundo? ¿Sí o no? El 99,99999 por ciento de la gente responde «no». Yo creo que tenemos que empezar a pensar más en cómo cambiar el mundo y en cómo nos puede ayudar a hacerlo el «alma de nuestros datos y la IA». Se nos presentan numerosas oportunidades para conseguir un gran impacto social, para transformar sociedades, para resolver problemas arraigados como la pobreza, el acceso a la educación o la escasez de comida, eliminando barreras y convicciones que ya no son ciertas (o que van desapareciendo). Así se allanará el camino a las nuevas oportunidades que pueden cambiar el mundo.

El futuro empieza ahora

Por si no lo sabías, si vuelas con frecuencia como yo, sin duda has sido alcanzado por algún rayo mientras volabas. La energía del rayo no atraviesa la cabina,

electrocutando a los pasajeros; nueve de cada diez veces se descarga al exterior, sin dejar apenas rastro de la descarga. La estructura del fuselaje y el potente aislamiento actúan como un pararrayos superconductor, canalizando el rayo por el exterior, lejos de los clientes y la tripulación y descargándolo en el suelo a través del tren de aterrizaje.

No obstante, el rayo deja en los aviones pequeñísimos agujeros que cuestan mucho dinero reparar, ya que requieren un meticuloso examen del avión, lo que retrasa la puesta en servicio del mismo. Los rayos que impactan en aviones suelen costar millones a las compañías aéreas, en detección y en reparación. Lo que suelen hacer es recurrir a técnicos expertos, entrenados para encontrar agujeros de 1 centímetro provocados por los rayos en determinados puntos del avión. No obstante, para ello hay que formar primero a estos expertos, ya que podrían pasar cosas por alto, o pueden cometer errores de identificación a causa de la fatiga, en particular cuando un avión llega a última hora de la noche y hay que ponerlo otra vez en servicio lo antes posible.

Actualmente, las aerolíneas han empezado a experimentar con drones que sobrevuelan el avión una vez ha aterrizado. Estos drones usan sofisticados mecanismos de captura de datos de imagen que se transmiten por internet a servidores en la nube para identificar los problemas usando un complejo software de inteligencia artificial. Los drones se comunican a través de la red y envían registros de imagen muy precisos con los que se pueden encontrar los rastros del impacto de los rayos rápidamente, para poder repararlos enseguida, y cometen muchos menos errores que los humanos.

Esto no es más que una mínima muestra de lo que nos espera en el futuro, un futuro que se acelera exponencialmente gracias a los grandes avances tecnológicos que van registrándose, y que cambiarán el mundo gracias a los datos y a la IA.

Los macrodatos y la IA se usan para analizar conjuntos de referencias tan grandes o complejos que no permiten el uso de las técnicas tradicionales de proceso de datos. Entre los desafíos que ello plantea está el análisis, la captura, el tratamiento de los anteriores, la investigación, la puesta en común, el almacenamiento, la transferencia, la visualización, la consulta, la actualización y la protección de la privacidad de la información.

Con el análisis de conjuntos de macrodatos se pueden encontrar nuevas correlaciones para detectar tendencias de negocio, para prevenir enfermedades, para combatir la delincuencia y mucho más. Ejecutivos, personal médico, publicistas y gobernantes se encuentran periódicamente con dificultades al tratar grandes conjuntos de registros en campos como la búsqueda en internet, las finanzas, la informática urbana o la informática de los negocios. Los conjuntos de datos crecen rápidamente en parte porque cada vez son más baratos y numerosos los detectores móviles, ficheros de registro, cámaras, lectores de identificación por radiofrecuencia (RFID) y redes de sensores *wireless*.

Hace diez años los científicos que analizaban los macrodatos eran una especie rara; hoy en día los datos y la IA son un activo de la mayoría de nuestras empresas y de la sociedad, lo que convierte a los analistas de *big data* y expertos en IA en los profesionales más solicitados en el mundo laboral. Y con el rápido avance de la inteligencia artificial es posible que pronto incluso muchos de ellos se vean reemplazados por máquinas que extraerán conclusiones automáticamente a partir de los datos y la IA. La velocidad a la que están cambiando las cosas es simplemente alucinante.

La capacidad de almacenaje de información mundial per cápita prácticamente se ha ido duplicando cada 40 meses desde los años 1980: en 2012, cada día se generaban 2,5 exabytes ($2,5 \times 10^{18}$) de datos. Sólo en el espacio de referencias de salud esperamos un aumento global del 48 por ciento anual. De hecho, en 2013 el volumen de datos sanitarios era de 153 exabytes. Con la tendencia de crecimiento prevista, para 2020 esa cifra aumentará hasta los 2.314 exabytes.

Para hacernos una idea, si tuviéramos que guardar todos éstos en un montón de tabletas, con las cifras de 2013 ese montón tendría casi 8.800 kilómetros de altura. Siete años más tarde, esa torre crecería hasta superar los 131.000 kilómetros de altura, llevándonos a más de un tercio de la distancia a la Luna.

Y la consecuencia de esto es que el exceso de datos, de macrodatos no procesados, puede tener el mismo impacto que un exceso de tráfico en una ciudad. Con el aumento del número de vehículos en las calles, las grandes áreas metropolitanas se van congestionando. Los atascos se convierten en un problema frecuente, y muchos trabajadores se enfrentan a la posibilidad de no poder llegar a tiempo a sus puestos de trabajo. Del mismo modo, un exceso de registros no procesados puede provocar una pérdida de oportunidades y puntos ciegos en nuestras empresas, o en nuestra vida diaria, que no nos permiten llegar donde queremos llegar.

Básicamente los macrodatos son datos que exceden la capacidad de procesamiento de los sistemas de datos convencionales. Éstos son demasiado grandes, se mueven demasiado rápido o superan las limitaciones de los ordenadores tradicionales. Para sacar un beneficio a estas referencias hay que escoger un modo alternativo de procesarlos.

El uso de *big data*, término de moda en 2012, se ha hecho viable gracias a la aparición de nuevos mecanismos capaces de «domar» el volumen, la velocidad y la variabilidad de estos datos enormes a un precio asequible. En el interior de éstos encontramos patrones e información de gran valor que antes quedaban fuera de nuestro alcance por la cantidad de trabajo necesaria para extraerlos. Las grandes corporaciones como Walmart o Google disponen de este poder desde hace tiempo, pero a un precio enorme. Hoy en día el hardware genérico, el acceso a la nube y el software de código abierto ponen el procesamiento de los macrodatos al alcance de los que menos recursos tienen. Ahora mismo pueden utilizarlos hasta una pequeña

empresa emergente creada en cualquier garaje, que por poco dinero puede comprar tiempo de proceso en la nube.

Llamamos macrodatos, o *big data*, a la siempre creciente capacidad que tenemos de recopilar referencias de multitud de fuentes y analizarlos para sacar conclusiones usando algoritmos informáticos avanzados. Hay patrones que los humanos no podemos ver, y que nos permiten comprender mejor las situaciones y las soluciones a los problemas. Los desastres naturales son situaciones tremendas, enrevesadas y complicadas, justo las condiciones perfectas en las que los macrodatos nos pueden ayudar a interpretar el caos. Las enormes cantidades de datos que generamos con los teléfonos móviles, los satélites y las redes sociales pueden ayudarnos a encontrar pistas sobre cómo responder a una situación de la mejor manera posible.

Gran parte del trabajo de desarrollo de los sistemas de *big data* para contribuir a aliviar los efectos de los desastres empezó tras el terremoto de Haití de 2010 y el terremoto de Tohoku (Japón) de 2011 y el posterior tsunami. Japón y Estados Unidos lanzaron un programa de investigación conjunto para reducir los efectos de los desastres naturales que matan a miles de personas cada año y que cuestan miles de millones a la economía global. El año pasado, la National Science Foundation de Estados Unidos y la Agencia de Ciencia y Tecnología japonesa ofrecieron 2 millones de dólares para financiar grupos que trabajaran con soluciones basadas en el uso de datos para afrontar la gestión de desastres.

En el otro extremo de la escala están las iniciativas de *crowdsourcing* creadas desde cero, en que los miembros de la comunidad se unen para ayudar a otros a través de la recopilación y el cotejo de registros. Esto es lo que ocurrió tras el paso del huracán Sandy por Estados Unidos, cuando la colaboración entre estudiantes de instituto permitió crear un mapa online de la zona de Nueva York y Nueva Jersey que mostrase dónde quedaba gasolina.

Tras el paso del tifón Haiyan por las Filipinas, la Cruz Roja internacional y UNICEF colaboraron con voluntarios de todo el mundo para trazar un mapa de los efectos sobre la región y su población. Los cuatro elementos clave de la gestión de desastres son la prevención, la preparación, la respuesta y la recuperación. Los macrodatos ofrecen la posibilidad de trabajar en los cuatro.

Existen muchos ejemplos más de situaciones y modos en que se pueden usar los macrodatos:

- Tu teléfono tiene acceso a un montón de resultados de macrodatos a través de *apps* que permiten orientarte con mapas, recibir recomendaciones musicales personalizadas de Spotify, comprobar el tiempo que hará mañana, hablar con Siri y otras muchas actividades diarias.

- El filtro de spam de tu gestor de correo es un ejemplo clásico de análisis de macrodatos para determinar qué es spam y qué no lo es, y luego aprende y adapta su inteligencia a tus necesidades al ir adquiriendo experiencia sobre tus preferencias personales. El termostato Nest hace lo mismo, ya que va adaptándose a tu rutina habitual y actúa en consecuencia.
- Cuando buscas un producto en Amazon y luego lo ves como producto «recomendado para ti» en otro sitio web, o cuando Facebook sabe de pronto a quién puede interesarte añadir a tu lista de amigos, están en acción los macrodatos y la inteligencia artificial, trabajando de forma combinada para que las webs intercambien información de lo que te gusta y luego usar esa información para decidir qué mostrarte. Lo mismo ocurre con la sugerencia de Amazon de «Los que han comprado esto también han comprado...».
- Cuando tu avión aterriza, no es un ser humano el que decide a qué puerta debe ir. Ni es una persona la que ha decidido el precio de tu billete; han sido una serie de cálculos con macrodatos.
- Los macrodatos y las búsquedas de internet pueden servir para predecir pandemias de gripe o prever la actividad de la gripe en todo el mundo. El objetivo es intentar hacer predicciones precisas sobre la actividad de la gripe, que pueden usarse posteriormente para crear modelos que prevean la extensión espacial y temporal de una enfermedad.
- El buscador de Google es un cerebro gigante con unos métodos increíblemente sofisticados para establecer una clasificación de las páginas y decidir qué enseñarte en particular. Lo mismo ocurre con el *feed* de Noticias de Facebook.

Y eso no son más que ejemplos del mundo del consumidor. Hoy en día se usan sofisticados sistemas de *big data* en numerosos sectores e industrias como las telecomunicaciones, la banca, los bufetes de abogados, el ejército, las fábricas y las finanzas (los operadores financieros algorítmicos de alta frecuencia mueven más de la mitad del capital social de los mercados de Estados Unidos), y en los sistemas de expertos como los que ayudan a los médicos a dar diagnósticos, o el famoso Watson de IBM, que contenía suficientes datos como para vencer a los grandes campeones del concurso de televisión *Jeopardy* sin problemas.

Pero los macrodatos también tienen limitaciones. «Te dicen lo que ocurrió en el pasado (o hace un momento), pero no siempre saben lo que ocurrirá en el futuro» (Mayer-Schönberger 2013). Así, se trate de la previsión meteorológica, de terremotos, del calentamiento global, de los resultados de fútbol, de los créditos *subprime* o de la crisis económica mundial, en muchos casos nos cuesta convertir los datos de ayer en predicciones por las que apostar. No obstante, sí podemos convertir el alma de éstos en nueva información con la que aprender y usar mejor nuestra intuición no para acertar siempre, pero sí para equivocarnos menos.

Un viaje al interior de los datos

La era de los macrodatos es una época en la que somos capaces de analizar las informaciones que proporcionan todo tipo de dispositivos conectados a internet (teléfonos, ordenadores, máquinas, coches, etc.) y convertirlos en predicciones y conocimiento procesable, lo que nos convierte en una «inteligencia aumentada». Este nuevo paradigma del conocimiento y el automatismo resultante se encuentran en el sustrato de la transformación económica que se avecina.

Todos estos cambios tecnológicos alterarán la existencia humana de un modo muy particular. Viviremos de un modo mucho más eficiente, y utilizaremos mejor los recursos gracias a los datos, pero sin dejar que ellos nos dominen a nosotros.

La historia de los macrodatos no es larga, pero gran parte de los cimientos sobre los que se apoya se pusieron hace mucho tiempo. Mucho antes de que los ordenadores (tal como los conocemos ahora) estuvieran tan extendidos, entre los estudiosos ya se fraguaba la idea de que estábamos creando un cuerpo de conocimientos cada vez mayor susceptible de ser analizado.

Quizá no siempre lo recordemos, pero nuestra creciente capacidad de almacenar y analizar información es producto de una evolución gradual —aunque desde luego las cosas se aceleraron a finales del siglo pasado, con la invención del almacenamiento digital e internet.

Ahora que sabemos que los macrodatos están destinados a ser objeto de uso común —tal como dijo Bernard Marr en el Foro Económico Mundial (*Breve historia del Big Data*, WeForum 2015)—, echemos un vistazo a la larga historia de pensamiento e innovación que nos ha llevado hasta el nacimiento de la era de los datos, según describe Bernard cronológicamente en dicho foro.

Historia antigua de los datos

- c. 18000 a. C.** Los primeros ejemplos que tenemos de almacenaje y análisis de datos por parte de seres humanos están en los palos de cómputo. El hueso Ishango fue descubierto en 1960 en la actual Uganda, y está considerado una de las primeras pruebas de almacenaje de éstos en la prehistoria. Las tribus paleolíticas hacían muescas en palos o huesos para registrar la actividad comercial o las provisiones. Luego comparaban palos y muescas para realizar cálculos rudimentarios, lo que les permitía hacer predicciones como cuánto les durarían las provisiones de alimentos.
- c. 2400 a. C.** Empieza a usarse en Babilonia el ábaco, primer dispositivo construido específicamente para realizar cálculos. Fue también en esta época cuando empezaron a aparecer las primeras bibliotecas, lo que representa nuestros primeros intentos de almacenaje de registros masivos.
- 300 a.C.-48 d. C.** La Biblioteca de Alejandría es quizá la mayor colección de datos del mundo antiguo: contendría quizá medio millón de pergaminos y todo el

conocimiento que habríamos adquirido hasta el momento, prácticamente sobre cualquier tema. Desgraciadamente, se cree que los romanos la destruyeron en el año 48 d.C., durante su invasión, quizá de forma accidental. En contra de lo que se suele pensar, no todo se perdió: una parte significativa de las colecciones de la biblioteca se trasladó a otros edificios de la ciudad, fue robada o se dispersó por el mundo antiguo.

- c. **100-200 d. C.** Creación del Mecanismo de Anticitera, el computador analógico más antiguo descubierto nunca, probablemente por científicos griegos. Su «CPU» está formada por 30 ruedas dentadas de bronce interconectadas y se cree que fue diseñado para realizar mediciones astrológicas y para predecir el ciclo de los Juegos Olímpicos. Su diseño sugiere que probablemente fuera una versión evolucionada de un mecanismo más antiguo, pero si es así, hasta ahora no se ha descubierto su precursor.

La emergencia de la estadística

- 1663** En Londres, John Graunt realiza el primer experimento de análisis estadístico de datos del que tenemos constancia. Registrando datos sobre mortalidad, teorizó que podía diseñar un sistema de advertencia precoz de la peste bubónica que asolaba Europa.
- 1865** Richard Millar Devens usa el término «inteligencia empresarial» en su *Enciclopedia de Anécdotas Comerciales y Empresariales*, al describir cómo adquirió ventaja sobre sus competidores, el banquero Henry Furnese, que recopiló y analizó de un modo estructurado la información relacionada con sus actividades empresariales, en lo que se cree es el primer caso de una empresa que aplica el análisis de informaciones con fines comerciales.
- 1880** La Oficina del Censo de Estados Unidos tiene un problema. Calcula que tardará ocho años en organizar todos los datos recogidos en el censo de 1880, y prevé que organizar los generados por el censo de 1890 llevará más de diez años. Eso significa que no estarán disponibles hasta después de la elaboración del censo de 1900, que lo habrá dejado obsoleto. En 1881, un joven ingeniero empleado en la Oficina, Herman Hollerith, crea lo que se dará en llamar el Tabulador Hollerith. Usando tarjetas perforadas, reduce el trabajo de diez años a tres meses y se hace un lugar en la historia como padre de la computación automatizada moderna. La empresa que funda más tarde se llamará IBM.

Los primeros días del almacenaje moderno de datos

- 1926** En una entrevista publicada en la revista *Colliers*, Nikola Tesla afirma que, cuando la tecnología sin hilos «se aplique perfectamente, toda la Tierra se convertirá en un enorme cerebro, que de hecho es, ya que todas las cosas son

partículas de un todo real y rítmico [...] y los instrumentos con los que podremos hacer esto serán sorprendentemente simples en comparación con nuestro teléfono actual. Podrán llevarse en el bolsillo del chaleco».

1928 Fritz Pfleumer, ingeniero germanoaustríaco, inventa un método para almacenar información magnéticamente en una cinta. Los principios que desarrolla aún se usan hoy en día, ya que la gran mayoría de datos digitales se almacenan magnéticamente en discos duros de ordenador.

1944 Fremont Rider, bibliotecario en la Wesleyan University de Connecticut (Estados Unidos), publica un artículo titulado «The Scholar and the Future of the Research Library» («El académico y el futuro de la biblioteca de investigación»).

En uno de los primeros intentos por cuantificar la cantidad de información que se produce, observa que para guardar todas las obras populares de valor que se crean, las bibliotecas estadounidenses tendrían que duplicar su capacidad cada dieciséis años. Eso le llevó a especular que la biblioteca de la Universidad de Yale, en 2040, contendría 200 millones de libros, distribuidos por 9.600 kilómetros de estanterías.

Los inicios de la inteligencia empresarial

1958 Hans Peter Luhn, investigador de IBM, define la inteligencia empresarial como «la capacidad de percibir las interrelaciones de los hechos presentados de modo que puedan guiar la actuación [de la empresa] hacia una meta deseada».

1962 Se dan los primeros pasos hacia el reconocimiento de voz, cuando William C. Dersch, ingeniero de la IBM, presenta la Shoebox Machine en la Feria Mundial de 1962. La Shoebox puede interpretar números y dieciséis palabras pronunciadas en inglés y convertirlas en información digital.

1964 Un artículo publicado en *New Statesman* hace referencia a la dificultad para gestionar la cada vez mayor cantidad de información de que disponemos.

La aparición de los grandes centros de datos

1965 El gobierno de Estados Unidos planea la creación del primer centro de datos del mundo para registrar 742 millones de devoluciones de impuestos y 175 millones de juegos de huellas digitales en cinta magnética.

1970 Edgar F. Codd, matemático de IBM, presenta su propuesta para una «base de datos relacional». El modelo proporciona el marco operativo que usan muchos servicios de registros modernos hoy en día, para archivar información en un formato jerárquico de modo que pueda acceder cualquiera que sepa lo que busca. Hasta este momento, para acceder a los datos de los bancos de memoria de un ordenador hacía falta la participación de un experto.

1989 Se usa, posiblemente por primera vez, el término macrodatos (*big data*) con el sentido que se le da hoy. Erik Larson, escritor superventas internacional, escribe un artículo en Harper's Magazine en el que especula sobre el origen del correo basura que recibe.

La aparición de internet

1991 El científico informático Tim Berners-Lee anuncia el nacimiento de lo que se convertiría en la Web tal como la conocemos hoy. En un post colgado en el grupo de Usenet alt.hypertext indica las especificaciones para una red mundial de datos interconectada, accesible para cualquier persona desde cualquier lugar.

1996 En su libro *The Evolution of Storage Systems*, de 2003, R.J.T. Morris y B.J. Truskowski afirman que este es el año en que el almacenaje digital empezó a ser más rentable que el almacenaje en papel.

1997 Michael Lesk publica su artículo «How Much Information is there in the World?» («¿Cuánta información hay en el mundo?»), teorizando que «quizá no sea exagerado suponer» una cifra de 12.000 petabytes. También señala que incluso en esta fase temprana de su desarrollo, la web va multiplicando su tamaño por diez cada año, y demuestra que gran parte de los datos nunca los verá nadie, por lo que no aportan ninguna información útil.

Google Search también hace su debut este año, y durante al menos dos décadas será sinónimo de búsqueda de registros en la red.

Las primeras ideas sobre macrodatos

1999 Apenas un par de años después, el término *big data* aparece en el artículo «Visually Exploring Gigabyte Datasets in Real Time» («Exploración visual de conjuntos de datos de gigabytes en tiempo real»), publicado por la Association for Computing Machinery. Una vez más, se lamenta la propensión a almacenar una cantidad significativa de referencias sin que haya posibilidad de analizarla adecuadamente. El artículo cita a Richard W. Hamming, pionero de la informática, que había dicho «El objetivo de la computación es la comprensión, no los números».

Posiblemente también sea el momento en que se usa por primera vez el término «Internet de las cosas» para describir el creciente número de mecanismos en línea y el potencial que tienen para comunicar unos con otros, en muchos casos sin un «intermediario» humano. El término se usa como título de una presentación realizada en Procter and Gamble por Kevin Ashton, pionero de la identificación por radiofrecuencia.

- 2000** Con el estudio «How Much Information?» («¿Cuánta información?»), Peter Lyman y Hal Varian intentan cuantificar por primera vez la cantidad de información digital del mundo, y su ritmo de crecimiento. Concluyen: «La producción total anual de contenidos en formato impreso, filmado, óptico y magnético del mundo requeriría aproximadamente 1.500 millones de gigabytes para su almacenaje. Eso equivale a 250 megabytes por persona, es decir por cada hombre, mujer y niño de la Tierra».
- 2001** En su estudio «3D Data Management Controlling Data Volume, Velocity, and Variety» («Gestión de datos en 3D, el control del volumen, la velocidad y la variedad de datos»), Doug Laney, analista de la Gartner Inc., define tres de las que acabarán siendo las características más comúnmente aceptadas de los macrodatos.

La web 2.0 aumenta los volúmenes de datos

- 2005** Los comentaristas anuncian que estamos presenciando el nacimiento de la «Web 2.0», la red generada por el usuario en la que la mayoría del contenido lo generarán los usuarios de los servicios, y no los proveedores de los mismos.

También es el año en que se crea Hadoop, la plataforma de código abierto creada específicamente para el almacenaje y el análisis de conjuntos de macrodatos. Su flexibilidad lo hace especialmente útil para gestionar registros desestructurados (voz, vídeo, texto sin formato, etc.) que generamos y recopilamos cada vez en mayor cantidad.

Se adopta el uso actual del término «big data» (macrodatos)

- 2007** La revista *Wired* acerca el concepto de los metadatos a las masas con su artículo «The End of Theory The Data Deluge Makes the Scientific Model Obsolete».
- 2008** Los servidores del mundo procesan 9,57 zettabytes (9,57 billones de gigabytes) de información, el equivalente a 12 gigabytes de información por persona y día), según el informe «How Much Information?» de 2010. En el artículo «International Production and Dissemination of Information», se calcula que este año se producirán 14,7 exabytes de nuevos datos.
- 2009** La empresa media con más de 1.000 empleados de Estados Unidos almacena más de 200 terabytes de informaciones, según el informe «Big Data The Next Frontier for Innovation, Competition, and Productivity» del McKinsey Global Institute.
- 2010** Eric Schmidt, presidente ejecutivo de Google, da un discurso en una conferencia en la que dice a los participantes que cada dos días creamos tantos datos como los que se han recogido desde el inicio de la civilización humana hasta el año 2003.

2011 El informe McKinsey advierte de que en 2018 Estados Unidos se enfrentará a una carencia de entre 140.000 y 190.000 profesionales expertos en datos y de que habrá que resolver problemas como la privacidad, la seguridad y la propiedad intelectual de ellos antes de que seamos conscientes del valor real de los macrodatos.

2014 De los ejecutivos que trabajan con Accenture, el 88 por ciento de los encuestados por GE afirman que el análisis de macrodatos está entre las prioridades de su empresa.

Lo que nos enseña Bernard es que los macrodatos no son un fenómeno nuevo o aislado, sino algo que forma parte de nuestra larga evolución de captura y uso de datos. Al igual que otros avances clave en el almacenamiento estos últimos, en los ordenadores y en internet, los macrodatos son un paso más que aportará cambios al modo en que gestionamos nuestras empresas y la red. Al mismo tiempo, sentará los cimientos sobre los que se forjarán muchos avances.

El McKinsey Global Institute (MGI) estudió los macrodatos en cinco dominios: la asistencia sanitaria en Estados Unidos, el sector público en Europa, las ventas al público en Estados Unidos y la manufactura y los datos de localización personal en todo el mundo. Los macrodatos pueden generar valor en todos ellos. Por ejemplo, un vendedor que use los macrodatos a fondo podría aumentar su margen de beneficios en más de un 60 por ciento.

Controlar grandes cantidades de información también puede suponer una gran ventaja en el sector público. Si el sistema de asistencia sanitaria de Estados Unidos usara los macrodatos de forma creativa y efectiva para potenciar la eficiencia y la calidad, el sector podría crear más de 300.000 millones de dólares en valor añadido cada año. Dos tercios de esa cantidad sería en la reducción del gasto sanitario nacional en un 8 por ciento aproximadamente. En las economías desarrolladas de Europa, los gestores gubernamentales podrían ahorrar más de 100.000 millones de euros en mejoras de la eficiencia operativa sólo con el uso de macrodatos, sin incluir el uso de las referencias para la reducción del fraude y de los errores y la mejora en la recaudación de impuestos.

Para que los datos sean útiles y se entiendan, deben ser recopilados y transportados a través de potentes redes de todo el mundo y, afortunadamente, también en ese campo hemos hecho grandes progresos. Por otra parte, éstos por sí solos no significan nada. Suele ser necesario establecer complejas relaciones entre ellos usando redes y conexiones que hay que analizar a fondo (por ejemplo, nuestras relaciones sociales en línea, los «me gusta» de nuestras páginas y de las páginas a las que llevan las páginas que nos gustan a nosotros, el patrón de actividad de telefonía móvil en una ciudad y cómo se mueve la multitud durante un evento, la red creada por los tuiteos al retuitearlos, etc.). Cuando observamos nuestros datos personales con

una visión de grupo, y analizamos cómo utilizarlos en la red, nos encontramos con cosas muy sorprendentes.

Si se utilizan los registros introducidos en la red, los «me gusta» de Facebook y los vínculos que has clicado, tu lista de reproducción de Spotify puede resultar muy útil. En lugar de tener que preguntarte si te gusta la poesía, ahora se puede echar un vistazo a tus «me gusta» de Facebook y obtener exactamente la misma información. Los datos y algoritmos que analizan cientos de miles de éstos pueden tener en cuenta pequeñas informaciones, sus interconexiones y los patrones de conexión en la red, combinarlos y predecir cosas muy complicadas. Si los datos están conectados, se pueden convertir datos aparentemente no relacionados en grandes historias sobre ti mismo y sobre tu potencial futuro.

No tiene nada de sorprendente echar un vistazo a tu lista de Spotify y ver que si tienes un montón de álbumes de Lady Gaga, quizá te guste ella y su tipo de música. Lo que resulta revolucionario del nuevo mundo de los datos y los algoritmos es que pueden usar tus preferencias musicales, o tus preferencias literarias, sus patrones de conexión en la red, y de esas relaciones aparentemente inocentes pueden hacer predicciones muy exactas de tu religiosidad, de tu rendimiento económico futuro, de los riesgos de tu empresa, de tu potencial de liderazgo, de tus opiniones políticas, etc. Esos modelos de relación de los registros en la red pueden generar mecanismos muy potentes para comprender el futuro, a las personas y la sociedad, y ofrecen unas oportunidades nunca vistas.

Google, el motor de búsqueda más potente del mundo, basa toda su inteligencia en uno de esos algoritmos de referencias en red, el PageRank, también conocido como el algoritmo algebraico lineal de los 25.000 millones de dólares. Este algoritmo es capaz de predecir cuáles son las páginas más importantes y cuáles deberían aparecer en primer lugar en el resultado de tus búsquedas, basándose en la red de datos creada a partir del modo en que están relacionadas unas páginas con otras, de qué sitios web dirigen a qué otros, y de lo importante que sean, lo cual se decide según quiénes sean los que los dirigen a ellos. El PageRank funciona contando el número y la calidad de los vínculos que llevan a una página para hacer un cálculo aproximado de la importancia de ese sitio web en relación con todo el conjunto de páginas de la red. Se supone que los sitios web más importantes probablemente tendrán más vínculos procedentes de otros sitios web. Las redes de datos y su estructura están en el centro de la revolución de esto últimos, el aprendizaje a partir de ellos y la inteligencia artificial.

El poder de las redes de datos

Barcelona, la ciudad donde vivo, no tiene puentes que crucen la ciudad. A diferencia de Königsbergin, población monástica prusiana que, tras quedar

prácticamente destruida durante la segunda guerra mundial y ocupada posteriormente por la Unión Soviética, fue rebautizada con el nombre de Kaliningrado. Fue fundada en 1255 por la Orden Teutónica y fue capital de la Prusia Oriental hasta 1945. Durante siglos fue un centro de aprendizaje donde vivieron Goldbach, Hilbert, Kant, y Wagner.

El río Pregel atravesaba la ciudad. Separaba dos grandes zonas en la costa y dos grandes islas. Había siete puentes que conectaban las diversas zonas, que dieron pie a un conocido problema matemático: los residentes de Königsberg se preguntaban si era posible recorrer toda la ciudad cruzando cada uno de los siete puentes una vez y no más.

Leonhard Euler (1707-1783), nacido en Basilea (Suiza), estaba destinado a convertirse en pastor religioso, como su padre. Estudió hebreo y teología en la universidad, pero en privado recibió clases de matemáticas. En 1727 consiguió un trabajo en el departamento de medicina de la Universidad de San Petersburgo, pero en el caos resultante tras la muerte de la emperatriz Catalina I, consiguió colarse en el departamento de matemáticas. En 1733 Euler estaba casado y tenía trece hijos, cinco de los cuales consiguieron sobrevivir hasta la edad adulta. En 1741 se trasladó a Berlín, donde vivió veinticinco años.

En toda su vida publicó más de 500 libros y artículos, y otros 400 se publicaron de forma póstuma. Cabe destacar que definió, entre otras, las notaciones i , p , e , \sin y \cos . Perdió la vista en ambos ojos, pero eso no hizo más que aumentar su productividad. «Ahora tengo menos distracciones», dijo, ¡y en 1736 hizo historia en las matemáticas al hablar de redes! (como internet, la web, las redes sociales, etc.).

Y lo hizo encontrando una solución al problema de los puentes de Königsberg (es decir, el de cómo pasar por todos los puentes sin pasar más de una vez por cada uno de ellos). Se dio cuenta de que no importaba cómo caminabas por el terreno, la forma de las islas, la longitud o la anchura de los puentes, ni dónde estaban los puentes exactamente. Sólo importaba cuántos puentes había entre cada pedazo de tierra, y en qué orden los cruzabas. Lo único que importaba era la estructura de red creada por los puentes. ¡Y fue un resultado espectacular!

La ciudad prusiana de Königsberg (ahora Kaliningrado, en Rusia) se encontraba a ambos lados del río Pregel y contaba con dos grandes islas, interconectadas entre sí y con la costa por siete puentes. El problema era idear un recorrido por la ciudad que supusiera cruzar cada puente una vez y sólo una, con estas condiciones: que a las islas sólo se podía llegar por los puentes, y que cada puente que se pisara se recorriera hasta el final. No era necesario finalizar el recorrido en el punto de partida.

Tras muchos largos paseos intentando encontrar la solución y mucho ensayo y error, Euler decidió idear una demostración matemática que dejara claro que el problema de Königsberg no tenía solución; nadie podía atravesar los puentes una sola vez y volver al punto de partida.