

Algoritmos, grandes datos e inteligencia en la red

Una visión crítica

Raymond Colle

Colección Mundo Digital

de Revista Mediterránea de Comunicación



9

Algoritmos, grandes datos e inteligencia en la red. Una visión crítica
Colección Mundo Digital de Revista Mediterránea de Comunicación

Directora

Victoria Tur-Viñes | Universidad de Alicante | victoria.tur@ua.es

Autor

Dr. Raymond Colle | Red I&I, Santiago de Chile | raymondcolle@gmail.com

Diseño y maquetación

Jesús Segarra-Saavedra | Universidad Internacional de La Rioja |
jesus.segarr@unir.net

Universidad de Alicante, España, 2017.

ISBN: 978-84-617-9283-2

doi.: 10.14198/MEDCOM/2017/9_cmd

Difunde: Grupo de investigación COMPUBES (Comunicación y Públicos Específicos)

Las fotografías incluídas en el texto provenientes de fuentes diversas se insertaron bajo el principio de "fair use", dado que la presente obra es de tipo académico y no tiene fines comerciales.

La cesión de derechos se realiza bajo la licencia Creative-Commons Reconocimiento Compartir Igual 3.0 (CC BY-SA 3.0 ES) de España.



* Queda expresamente autorizada la reproducción total o parcial de los textos publicados en este libro, en cualquier formato o soporte imaginable, salvo por explícita voluntad en contra del autor o autora o en caso de ediciones con ánimo de lucro, señalando siempre la fuente. Las publicaciones donde se incluyan textos de esta publicación serán ediciones no comerciales y han de estar igualmente acogidas a Creative Commons. Harán constar esta licencia y el carácter no venal de la publicación.

Índice

	Introducción	5
1	Definiciones iniciales	6
2	Grandes datos (<i>Big Data</i>)	7
3	Algoritmos	19
4	Inteligencia artificial	27
5	Después de los <i>Big Data</i>	32
6	<i>Big data</i> en la prensa	35
7	Poder y ética en la era de los datos	38
8	¿Una internet inteligente?	53
	Conclusión	56
	Bibliografía	58
	Notas	59

Introducción: La acumulación de datos personales

"Al entregar los datos a la red ya no nos poseemos a nosotros mismos"
(Peter Hirshberg, en El País.es)

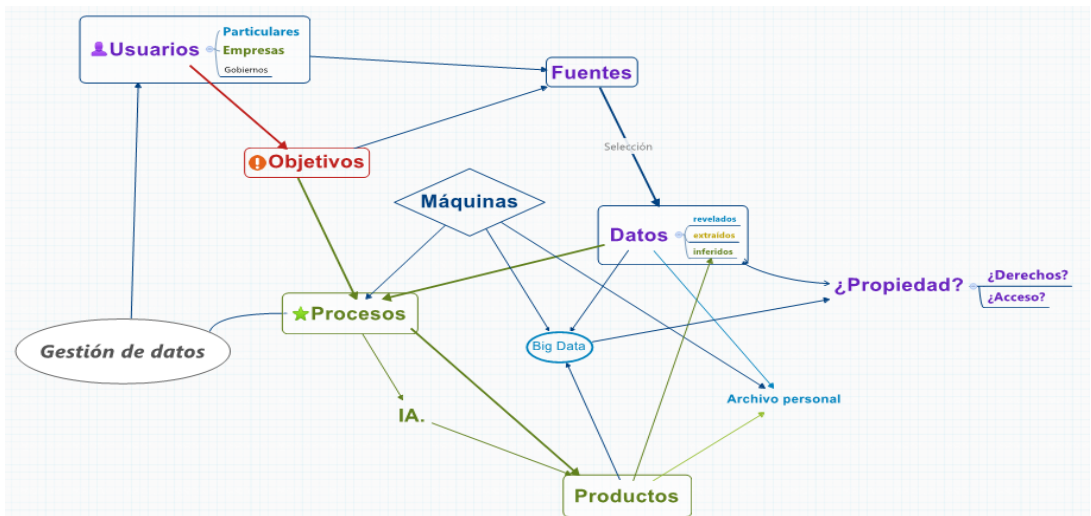
"Existimos en diversos servidores como una enorme colección de datos."
(Ángel Jiménez de Luis, en El Mundo.es)

Al utilizar la web, los teléfonos y los televisores "inteligentes", proporcionamos informaciones sobre nosotros mismos, que lo queramos o no. Siempre dejamos rastros de nuestra actividad en la web, sea que usemos las redes sociales o no. Y todas las empresas ahí presentes tratan de capturar esta información, que es ahora un valioso producto comercial. Y éste crecerá exponencialmente con los *gadgets* que formarán la llamada "Internet de las Cosas" o IoT (*Internet of Things*). Nuestras actividades *online* son fuente de conocimiento no solo para nosotros mismos sino para quienes nos prestan sus servicios en la red.

Los datos recopilados en internet por las empresas se acumulan en grandes bases de datos que han pasado a llamarse "*big data*". Google procesa 3.500 millones de búsquedas por día y almacena unos 10 exabytes¹ de datos. Tiene más de un millón de servidores. Facebook agrega 500 terabytes de datos diariamente. Amazon mantiene los datos de compras de 152 millones de clientes en alrededor de 1 exabyte. (Fuente: Adeptia; Infographic Journal, 6/03/2015).

Los captores de datos personales, en el mundo de internet, no son todos equivalentes ni tienen los mismos objetivos. Las empresas tratan de aprovechar los contactos logrados a través de la web o de sus *apps* para conocer mejor a sus posibles clientes y rentabilizar mejor la relación. Los gobiernos ofrecen y recopilan información para orientar su gestión y mejorar sus servicios... y realizan actividades de espionaje para proteger su seguridad. Los políticos están interesados en la opinión pública y en la forma de influir en ella. Y no olvidemos los piratas (*hackers*) que tratan de acceder a informaciones reservadas, servicios financieros, infraestructuras, etc. con las intenciones más oscuras. Las instituciones académicas, además de darse a conocer, ofrecen cada vez más alternativas de enseñanza a distancia y publican los resultados de sus proyectos de investigación.

El siguiente esquema muestra la complejidad del tema de la gestión de datos privados en el ambiente actual. Abordo más detalladamente este tema en mi libro "*Secuestradores y secuestrados: Nuestros datos en internet*".



1. Definiciones iniciales

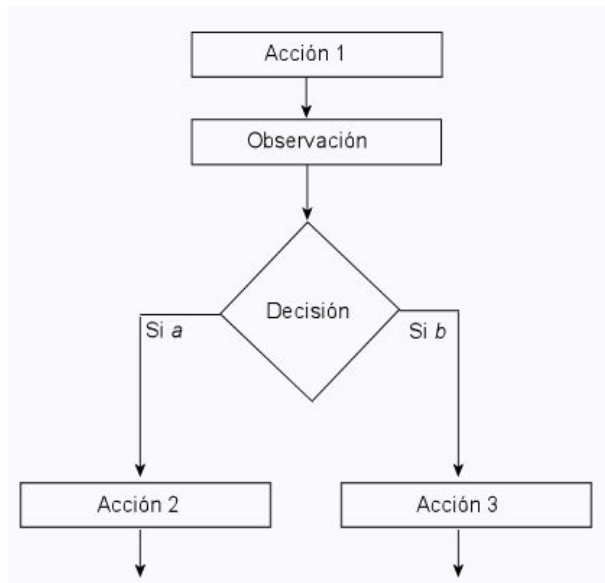
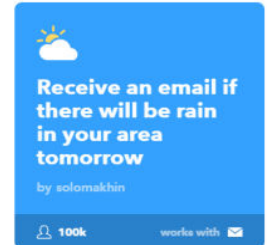
1.1. Algoritmos

La Wikipedia nos da una muy buena explicación de lo que es un algoritmo:

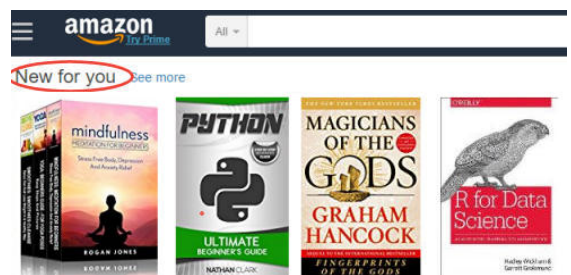
Un algoritmo (del griego y latín, dixit algorithmus y este a su vez del matemático persa Al-Juarismi) es un conjunto prescrito de instrucciones o reglas bien definidas, ordenadas y finitas que permite llevar a cabo una actividad mediante pasos sucesivos que no generen dudas a quien deba hacer dicha actividad. Dados un estado inicial y una entrada, siguiendo los pasos sucesivos se llega a un estado final y se obtiene una solución.

Un algoritmo podría ser una simple secuencia de órdenes: "haga esta acción, luego esta otra, luego la siguiente..." pero la verdadera "magia" de un algoritmo reside en la opción condicional: "si ocurre esto, haga tal cosa; si ocurre esta otra, haga tal otra cosa". Es lo que ilustra en forma espléndida el servicio web IFTTT (<https://ifttt.com/>): "if this then that", además con sus apps para teléfonos iPhone y Android.

Los algoritmos pueden ser programados de diferentes formas, según el lenguaje de programación utilizado, pero la mejor forma de explicar su lógica (y de preparar la programación) es la forma gráfica. Aquí el modelo más básico, con su geometría estándar:



Los algoritmos se utilizan hoy en forma intensiva en muchos sitios web comerciales (como en Amazon) para ofrecer productos o servicios a partir de la detección de las acciones anteriores de los usuarios (que permiten suponer preferencias) y en las redes sociales (también para hacer ofertas y, sobretodo, para acumular conocimiento propio acerca de los usuarios (y eventualmente vender este). Los buscadores, como Google, recurren a algoritmos altamente complejos (con decenas de condiciones) para



estructurar sus páginas de resultados en torno a lo que podría responder "mejor" a los intereses de quien busca. Esta necesidad de considerar múltiples condiciones es la que complica el análisis de grandes cantidades de datos (*big data*) y que lleva a la utilización de una inteligencia artificial.

1.2. Grandes datos (Big Data)

Todos los datos recopilados por las empresas, servicios y redes sociales que "visitamos" en internet son reunidos en enormes bases de datos, los llamados "big data", donde pueden ser analizados para proporcionar más conocimientos acerca de nosotros.

Se acumulan tantos datos que ya no es necesario que nuestro nombre esté registrado en la base de datos – y tampoco un perfil de usuario – para que seamos identificados. Nuestra identidad puede ser determinada sobre la base de nuestro comportamiento en algunas aplicaciones como compras y pagos online, servicios multimedia y plataformas de comunicación aún sin nuestro nombre?

¿Que datos?

Los datos personales que pueden aparecer y son recopilados cuando utilizamos internet son de cuatro tipos:

- privados: son los que generamos y que "viajan" por medio de sistemas directos de persona a persona (como el correo electrónico y el chat); también pueden incluir documentos en la nube (Dropbox, iCloud, etc.) que no compartimos;
- públicos: son nuestras publicaciones en la web: páginas web personales, blogs, artículos de revistas, documentos en servicios como ISSUU, Scribd, etc.
- sociales: corresponden a nuestras actividades en las redes sociales (mensajes de estado, fotos y vídeos, "likes", etc.);
- "rastros" (data trail): huellas dejadas involuntariamente por las actividades individuales (p.ej. datos del aparato utilizado, ubicación, enlaces seguidos, etc.).

Los datos públicos y sociales son información revelada voluntariamente. Los datos privados pueden ser analizados con o sin nuestro consentimiento si no son encriptados y dejan siempre algunos rastros utilizables por los proveedores de servicios. Constituyen información "extraída" de nuestro comportamiento en las redes.

1.3. Inteligencia artificial

1.3.1. Origen y pretensiones

Ante todo, ¿que significa realmente "inteligencia artificial"? El concepto es fruto de unas famosas conferencias que tuvieron lugar en 1956 en el Dartmouth College, en Hanover (New Hampshire, Estados Unidos). Los investigadores reunidos ahí discutieron acerca de la posibilidad de construir máquinas que no se limitaran a hacer cálculos prefijados sino operaciones genuinamente "inteligentes". Y, para ello, se inspiraban en los trabajos de Alan Turing quien, deseoso de crear un cerebro artificial, inventó la siguiente prueba:

Un interrogador permanece encerrado en una pieza, enviando mensajes por un buzón a dos interlocutores, un hombre y una mujer, siendo su objetivo determinar quién es el hombre y quién la mujer (sin preguntar, obviamente, por el sexo de cada uno). En el modelo inicial de la prueba, se supone que el varón quiere hacerse pasar por mujer. Luego se pregunta: ¿qué ocurrirá si el varón es reemplazado por una máquina? Si el diálogo que ocurra y el número de errores en la solución dada se acerca al número de errores ocurridos en la comunicación con un ser humano, se podrá estimar –según Turing– que estamos ante una máquina "inteligente". (Tengamos en cuenta que la diferencia de sexo era mucho más fácil de descubrir en su época –1950–, razón por la cual este aspecto desapareció en las aplicaciones y en muchos comentarios posteriores de este test. Esto significa que las exigencias –para equiparar la máquina y el sujeto que trata de engañar– eran bastante modestas).

A partir del grupo del Dartmouth College, se formaron dos grandes "escuelas" de IA: la de la Universidad de Carnegie-Mellon que se propuso desarrollar modelos de comportamiento humano con aparatos cuya estructura se pareciese lo más posible a la del cerebro (lo que posteriormente derivó en los trabajos sobre redes neuronales artificiales) y la del Instituto Tecnológico de Massachusetts (MIT), que se centró más en que los productos del procesamiento tuvieran el carácter de inteligente, sin preocuparse por el funcionamiento o la estructura de los componentes (lo cual se podría eventualmente lograr con computadores normales).

En la propuesta de Turing – y en las fórmulas mejoradas del test que se han construido después – es importante tener en cuenta que la definición de la inteligencia está íntimamente relacionada con el protocolo que se estima correcto para determinar su existencia. Desde este punto de vista, Turing habría concebido el test de tal modo que fuese fácil considerar a los computadores como "inteligentes", ya que se "define" la inteligencia como lo que el computador logra demostrar en dicho test. Pero es una definición muy peculiar (y limitada) de la inteligencia, como han recalcado los psicólogos expertos en desarrollo de la inteligencia, los cuales han aportado múltiples argumentos que tienden a situar la inteligencia fuera del alcance del protocolo de Turing. Suponiendo que una máquina pasa con éxito el test de Turing (incluso con mayores exigencias), estaremos frente a un ente que reacciona adecuadamente, pero ello no implica que tenga entendimiento (o que sea realmente "inteligente" al modo humano) (H. Collins).

El análisis del lenguaje fue una de las áreas en que más se trabajó. Se descubrió pronto que el mero reconocimiento del léxico no llevaba a ninguna parte y los trabajos de Chomsky llevaron a estudiar cómo era posible convertir una cadena lingüística – con una estructura determinada – en otra, lo cual sería clave para los sistemas actuales (aunque se ha renunciado a aplicar directamente sus ideas). Se llegó así a reconocer que la comprensión del lenguaje se demuestra fundamentalmente en la capacidad de parafrasear, es decir de referirse a un mismo contenido cambiando su forma de expresión (cfr. Schank, pp.108-109). Dice Sánchez-Migallón:

Aunque se han dado ciertos avances, no se ha solucionado el problema del sentido común y las computadoras siguen sin poder enfrentarse a entornos poco formalizados. Hemos descubierto que la inteligencia no se reduce a mero cálculo (en lo cual las máquinas son las campeonas), sino que hay muchos tipos de inteligencias muy interconectadas: hay inteligencia social, emocional, creativa... ¿Cómo implementar en una máquina algo parecido a una emoción? No hay ingeniero en el mundo que tenga la más remota idea de cómo hacerlo.

¡Aunque ya se aplican programas de inteligencia artificial para detectar e identificar emociones!

1.3.2. Inteligencia de máquina

Las referencias a los *big data* y al análisis de los mismos remiten casi siempre al uso de procedimientos de inteligencia artificial para extraer información. Pero es necesario introducir algunas precisiones en torno a este concepto.

No es difícil definir "artificial" ya que, simplemente, es algo que no existe en la naturaleza. Pero definir "inteligencia" puede ser un problema porque no todos los filósofos o psicólogos entregan la misma explicación. El mismo David Chalmers, un experto en el tema, sugirió incluso el año 2015 que no se debía esperar unanimidad siquiera en este siglo. Y John McCarthy, un pionero de la IA en la Universidad de Stanford abordó la cuestión en noviembre de 2007:

"Q: ¿No hay una definición sólida de inteligencia que no dependa de relacionarla con la inteligencia humana? R: Todavía no. El problema es que todavía no podemos caracterizar en general qué tipo de procedimientos computacionales queremos llamar inteligente. Entendemos algunos de los mecanismos de la inteligencia y no a otros." (Agenda World Economic Forum, 7/09/2015).

A pesar de ello, Marcus Hutter y Shane Legg han propuesto la siguiente definición general: "*la inteligencia mide la capacidad de un agente para lograr objetivos en una amplia gama de entornos*" (*ibidem*). Pero podría ser demasiado amplia (¿no hay muchos programas de computación que logran esto pero que nadie consideraría "inteligentes"?) y es dudoso que consigan una aceptación unánime, especialmente por parte de los psicólogos.

Entre los últimos avances en materia de "inteligencia de máquina" y "aprendizaje automático", se puede señalar lo logrado por los ingenieros del Massachusetts Institute of Technology: su IA es capaz de imitar al ser humano para hacer una búsqueda en la red, incluso refinándola si la respuesta inicial no parece satisfactoria... aunque los administradores deben intervenir al principio. "*El sistema, según explica Regina Barzilay, líder del proyecto, fue creado para clasificar automáticamente datos de texto, combinando todo lo recopilado, para encontrar patrones que correspondan a diferentes categorías proporcionadas por los propios administradores de la plataforma*" (FayerWayer, 11/11/2016).

1.3.3. Sistemas expertos

Aunque se habla de las interfaces de voz como Siri y Cortana como sistemas de IA, en realidad son solo sistemas expertos, que corresponden a una fase inicial de desarrollo de la IA (por esta razón implantable hoy en aparatos como los *smartphones*, ya que utilizan menos recursos). Contienen información básica que les permiten realizar operaciones en un dominio muy específico y, actualmente, pueden además "aprender" a partir del uso que se les da (básicamente a partir de una estadística de los términos dictados). Sin embargo, solo han sido posibles gracias a trabajos previos de poderosos computadores con procesadores que funcionan en forma paralela y que aprenden patrones a partir de enormes volúmenes de información. Este trabajo ha sido mérito de Google, que reúne desde 2008 en sus computadores la información entregada por sus usuarios, logrando llevar la cifra de aciertos, de acuerdo a sus investigaciones, al 90% de exactitud en el reconocimiento del habla humana (BBC Mundo, 15/09/2015).

2. Grandes datos (*Big Data*)

2.1. ¿Para qué?

Las empresas de internet – como muchas otras – acumulan una cantidad enorme de informaciones, superior a lo que es posible humanamente conocer y manejar. Facebook, Google y otros realizan un seguimiento permanente de sus usuarios y acumulan estas informaciones. Incluso los medios de prensa pueden conocer los intereses de sus lectores. Son cantidades ingentes de datos, que conforman conjuntos gigantescos llamados por ello "*big data*". Con los móviles y las redes sociales, las empresas conocen nuestras listas de amigos, nuestros gustos, donde hemos estado, y más.

Si bien declaran utilizar los "rastros" de nuestra actividad en sus servicios, gigantes como Google, Facebook y Twitter los combinan con todo lo que les hemos dicho públicamente de nosotros mismos y lo que han observado sin que lo sepamos no solo para "mejorar su servicio" sino para obtener ingresos mediante la venta de estos datos. Así,

En vez de lanzar un mensaje al aire basándose en estimaciones de la audiencia, los anunciantes ahora pueden segmentar sus comunicaciones con un nivel de detalle impensable hace sólo una década.

Si, por ejemplo, quiero que mi anuncio llegue sólo a jóvenes solteros del barrio de Salamanca en Madrid, ahora puedo hacerlo. Si prefiero mujeres de más de 60 años con nietos en Bilbao, también. Mi restaurante puede mostrar publicidad en los navegadores de aquellos que están en los códigos postales cercanos o mi tienda de ropa anunciar las últimas novedades junto a los resultados de búsqueda en aquellos hogares que generen más de 80.000 euros en ingresos.

Para Facebook esto se traduce en que cada usuario de la red social vale cerca de 1,60 euros por trimestre. Twitter puede llegar a exprimir cientos de euros al año en valor de sus perfiles más activos. No son cantidades sorprendentes, pero al multiplicar por los cientos de millones de usuarios de estos servicios, las cuentas salen redondas (El Mundo.es, 20/11/2014).

A pesar de que estos datos, por lo general, están "anonimizados", ya sabemos que cruzando diferentes puntos es sencillo obtener un perfil único y real de cualquier usuario. No hay forma de que nos escondamos, a menos que dejemos de usar la red. Por esto, nuestra identidad y nuestros datos son el precio que hemos de pagar por usarla.

¿Cuáles son los usos que las empresas dan a las herramientas de análisis de datos? Realizan principalmente (48%) un análisis del comportamiento de los consumidores para poder predecir el comportamiento de los consumidores (Betanews, 21/11/2014). Este es uno de los campos que más interesan y se espera que el marketing predictivo personalizado, utilizando los datos sociales, sea una de las áreas de negocio que más se beneficiarán del análisis de *big data* en los próximos años, a pesar de que el 71% de los directores de marketing de todo el mundo dicen que su organización no está preparada para hacer frente a esta explosión de datos en los próximos 3 a 5 años (Business Insider, 12/05/2014). Y es que analizar estas montañas de datos de modo que se obtenga información útil no es tarea fácil, como lo muestra el hecho de que los grandes "recopiladores" intenten recurrir a la inteligencia artificial, como ocurre con Facebook, Google, Twitter, LinkedIn y algunos otros. Ellos están comenzando a utilizar estas técnicas avanzadas para desarrollar su capacidad de "aprendizaje profundo" a partir de los datos que acumulan en sus redes, desde las conversaciones hasta el reconocimiento facial de las fotos y la actividad en los juegos. Así, tienen el potencial de ser mucho más

personalizados. Y hacen emerger nuevos campos de marketing: la agrupación de audiencia (*clustering*), el marketing predictivo y el análisis de los sentimientos frente a las marcas.

Analizando o vendiendo estos datos, las empresas pueden obtener mayores beneficios gracias a la información que inocentemente les entregamos. Alex "Sandy" Pentland, un científico de datos, afirma que la capacidad de recoger estos datos sobre nuestro comportamiento permitirá a los científicos desarrollar "*una teoría causal de la estructura social*" y, finalmente, establecer "*una explicación matemática de por qué la sociedad reacciona como lo hace*" en todo tipo de circunstancias... e incluso predecir comportamientos.

La cantidad de aplicaciones y usos posibles no ha dejado indiferente a la Comisión Europea. Ha pedido a los gobiernos nacionales que "*abran los ojos ante la revolución del Big Data*" y, además de establecer una serie de centros de supercomputación de excelencia y crear una incubadora de datos abiertos, ha propuesto realizar una cartografía de normas sobre datos, identificando las posibles lagunas y proponiendo nuevas reglas en lo referente a la propiedad de los datos y a la responsabilidad del suministro de los mismos (TICbeat, 5/07/2014).

Uno de los temas más delicados es lo que ocurre con los datos relacionados con la salud, un problema que crece al mismo tiempo que la multiplicación de *apps* relacionados con el estado físico (*fitness*) y del uso de buscadores por pacientes que quieren saber más acerca de sus malestares. Al respecto, un investigador de la universidad de Pennsylvania, Tim Libert, creó un *software* para analizar páginas web y saber hacia dónde iban las búsquedas y lo aplicó a páginas de medicina. Descubrió que el 91% de las páginas hacían llamadas a terceros cuando se realizaba una búsqueda. Es decir que cuando se busca información sobre una enfermedad y se pincha en un enlace, es muy probable que esos datos se estén compartiendo con otras empresas. Según Libert, la firma Experian (de comercio de datos) está en el 5% de las páginas relacionadas con medicina pero no se sabe qué datos almacenan ni cómo lo hacen (Xataka, 24/02/2015).

2.2. Quiénes

Facebook, Twitter, LinkedIn y otros están comenzando a utilizar técnicas de inteligencia artificial para desarrollar su capacidad de aprendizaje "profundo" a partir de los datos que acumulan a través de sus redes, desde las conversaciones hasta el reconocimiento facial de las fotos y la actividad en los juegos. Así, tienen el potencial de ser mucho más personalizados. Y hacen emerger nuevos campos de marketing: la agrupación de audiencia (*clustering*), el marketing predictivo y el análisis de los sentimientos frente a las marcas. Facebook dispone para ello de un laboratorio de investigación dedicado a la inteligencia artificial (IA); Google adquirió DeepMind, una compañía que agrupa los mejores talentos en IA y crea algoritmos de análisis para el e-comercio; LinkedIn compró Bright, una compañía parecida, que desarrolla algoritmos de selección de trabajos; Pinterest adquirió VisualGraph, que reconoce imágenes. (Wired, 24/04/2014).

Greg DeMichillie, director de gestión de productos de Google en la "nube", anunció que, en 2015, la empresa se centraría en el lanzamiento de herramientas y servicios que facilitarían las operaciones con grandes datos. Será "Google Cloud Dataflow", un servicio gestionado de procesamiento de datos, diseñado para crear canales de datos que "*ingieran, transformen y analicen los datos, tanto en lotes como en flujos transmitidos*" (ComputerWorld, 26/06/2014). Pero van más lejos: la totalidad de nuestra información genética podría ser almacenada en la nube gracias a Google Genomics. Tan sólo el genoma de una persona puede alcanzar los 100 GB de información. Google ya ha logrado conseguir la información genética de 3.500 personas y la almacena en los servidores de la compañía. Almacenar el genoma cuesta USD \$25 al año (FayerWayer, 7/11/2014).

Microsoft ha lanzado un servicio en la nube llamado "Azure Machine Learning" (AzureML), un sistema de aprendizaje automático que permite analizar los *big data* para reconocer patrones y extraer valor. Ofrece una interfaz tipo web, de arrastrar y soltar, para colocar fácilmente las piezas del rompecabezas y extraer información sobre la base de conjuntos de datos complejos (Evenbrite.ca, 13/11/2014).

IBM está invirtiendo miles de millones de dólares en su división de investigación dedicada a lo que llaman "*computación cognitiva*", un término que la compañía utiliza para referirse a las técnicas de inteligencia artificial relacionadas con su supercomputador Watson. Los esfuerzos de investigación son de largo alcance, e incluyen el desarrollo de componentes electrónicos conocidos como chips neurosinápticos, que tienen características modeladas en base al funcionamiento de los cerebros biológicos y son más eficientes en el procesamiento de información sensorial. A pesar de ello, algunas empresas e investigadores que han estado probando Watson señalan haber tenido dificultades para adaptar la

tecnología y hacer que funcione con sus conjuntos de datos. "No está despegando tan rápido como les gustaría", señaló el profesor de administración en la Escuela de Negocios de Copenhagen, Robert Austin, que ha estudiado la estrategia de IBM durante los últimos años. "Esta es una de esas áreas en las que hacer que una demostración se convierta en algo de valor empresarial real depende de detalles poco conocidos." (TechnologyReview.es, 7/11/2014).

Y es que la inteligencia artificial, si bien tiene unos 50 años de historia, aún tiene muy poco de inteligente. Aunque los sistemas más avanzados utilizan métodos inspirados en la estructura del cerebro (como las redes neuronales artificiales, simuladas en supercomputadores, que facilitan el reconocimiento de patrones) o sus procedimientos (como el reconocimiento de secuencias de operaciones), estamos lejos de acercarnos a la capacidad del cerebro humano (con sus 100 billones de neuronas, 10^{15} conexiones y 10^{27} operaciones por segundo y por neurona).

Más vale no hacerse ilusiones acerca de sus avances. Empresas como Netflix pueden extraer tendencias y ofrecer contenidos cercanos a los intereses de sus clientes (como también pueden hacerlo los medios de prensa), pero no se requiere inteligencia artificial para ello: solo un buen sistema de clasificación y una aplicación estadística básica.

2.3. Pretensiones empresariales

Empresas como Google, Facebook, Microsoft y otras han contratado los mejores expertos en inteligencia artificial para sacar partido de todos los datos que acumulan. En los últimos cinco años, los avances en inteligencia artificial, en particular en la rama de las redes neuronales, han generado productos de avanzada que ya están llegando a nuestras vidas a través de los laboratorios de IA de empresas como Google, Facebook, Microsoft y Baidu, que están colocando a cientos de millones de dólares en la carrera por mejores algoritmos y computadores "más inteligentes".

El "aprendizaje profundo" de los nuevos sistemas de IA ha impulsado el reconocimiento de voz en Apple y Android, y ya permiten la traducción instantánea en Skype (chats de texto en más de 40 idiomas, y traducción hablada en tiempo real entre hablantes de español e inglés). Google construye autos que se conducen solos. Los sistemas informáticos pueden enseñarse a sí mismos para identificar objetos o personas en fotografías.

De ahí el convencimiento tanto de Elon Musk, Stephen Hawking y varios otros científicos de que ha llegado el momento de plantearse la problemática ética ligada a estos avances, tal como lo hicieron los biólogos moleculares en 1975, en la Conferencia de Asilomar sobre ADN Recombinante, donde definieron las normas de seguridad necesarias para prevenir que organismos artificiales modificados genéticamente pudieran representar una amenaza para el público (ver apartado 7.4.1.).

En diciembre de 2014, Eric Horvitz (de la Universidad de Stanford) inició un estudio para analizar durante 100 años los efectos de la evolución de la inteligencia artificial. Y el Allen Institute for Artificial Intelligence, impulsado por Paul Allen (cofundador de Microsoft), lanzará un programa piloto de inteligencia artificial llamado *Semantic Scholar*, que ayudará a los investigadores a gestionar la constante avalancha de nuevas publicaciones académicas (TICbeat, 17/01/2015).

Y no se ha de pensar que la IA sigue reservada a las grandes empresas o universidades y que éstas nos "servirán" a través de programas que operen en su "nube" (como en el caso de Skype Translate). Spark ya ofrece una consola de inteligencia artificial, la *EmoSpark*, que se conecta vía *WiFi* o *bluetooth* al televisor y computadores del hogar y a la que se podrá pedir que haga una llamada telefónica o que busque cuál es la capital de algún país. "También detecta el estado de ánimo, ayuda a navegar por la red, notifica actualizaciones y mensajes en tus redes sociales y graba todo lo que te gusta para recordártelo cuando estés triste." Todo ésto por USD \$315 (El Mercurio 3/01/2014).

Una vez más, nos hacen pensar que "la Singularidad está cerca" (el momento en que la IA podría superar la mente humana), como advierte Ray Kurzweil (aunque "cerca" quiere decir hacia la segunda mitad de este siglo y que, posiblemente, sea demasiado "optimista" al respecto).

2.4. Herramientas

2.4.1. Minería de datos

El concepto de minería de datos o *data mining* agrupa técnicas computacionales que permiten descubrir informaciones poco visibles en las bases de datos, especialmente características que relacionan de un modo inesperado – o difícil de descubrir – los valores de múltiples variables en una gran cantidad de registros.

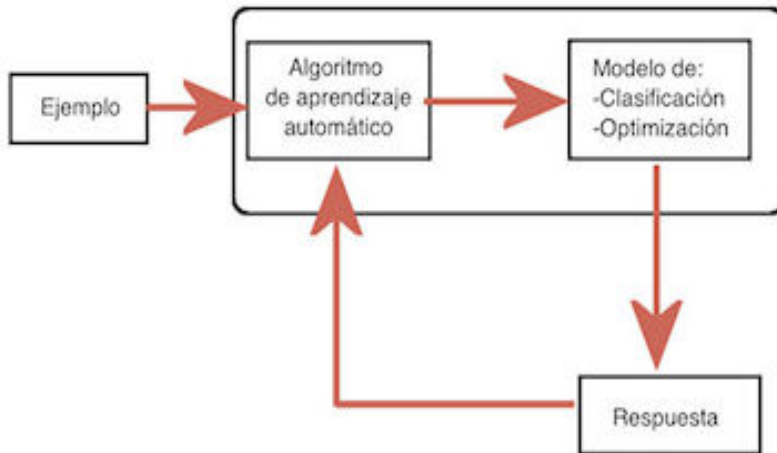
Los métodos de DM (*data mining*) permiten descubrir esta información y transformarla en un valioso conocimiento tanto retrospectivo (histórico) como prospectivo (proyecciones) o “comprensivo” (entender lo que ocurre), siendo así muy importante para las tomas de decisiones en las empresas, organizaciones y gobiernos. Por lo tanto, la DM es por esencia una metodología de exploración y descubrimiento. Una vez obtenido un resultado – por ejemplo un patrón de comportamiento de un posible sujeto (natural o jurídico) – y transformado dicho resultado en modelo de contrastación para ser utilizado repetidamente con posterioridad, este nuevo uso – que también es una investigación – ya no puede ser llamado “minería de datos”, porque no apunta a descubrir un conocimiento nuevo, aunque puede formar parte de los nuevos métodos de administración basados en el análisis de *big data*. (Westphal y Blaxton:6)

La minería de datos no elimina la participación humana para resolver la tarea completamente pero simplifica significativamente el trabajo y permite a un analista que no es un profesional de las estadísticas y de la programación administrar el proceso de extraer conocimiento a partir de los datos. Los métodos de DM son variados y el *software* existente incluye habitualmente una “batería” de programas que operan de distintas maneras y entregan distintos tipos de resultados, en su mayoría acompañados de formas visuales destinadas a poner en evidencia las relaciones descubiertas. Las operaciones que es posible realizar son, por ejemplo, la clasificación, la estimación, la predicción, la detección de relaciones, el modelamiento, la segmentación (*clustering*) y la detección de desviaciones.

La DM ha sido utilizada para investigar fraudes y el crimen organizado, en análisis financiero (bancos, bolsas), telecomunicaciones (operación de redes), medicina (influencias de la interacción de medicamentos), etc. El motor de búsqueda de Google es un ejemplo de aplicación de este tipo, pero basado en un algoritmo altamente complejo. (He analizado detalladamente esta técnica en mi libro “*Explotar la información noticiosa*”).

2.4.2. Aplicaciones de IA

El “análisis de *big data*” se basa en los métodos de la minería de datos pero los amplía recurriendo a técnicas propias de la inteligencia artificial (o, más bien, de “inteligencia de máquina”), especialmente la capacidad de inducir conocimiento a partir de la acumulación de información no estructurada suministrada en forma de ejemplos o casos (La DM requiere datos estructurados). Esta acumulación se complementa con los datos de las mismas operaciones realizadas y cada resultado positivo de las búsquedas refuerza los circuitos (algoritmos) que facilitan la obtención de mejores respuestas. De este modo, la programación apunta a generalizar comportamientos o diseñar soluciones factibles a diversos problemas específicos. Es el llamado “aprendizaje automático” o “*deep learning*”.

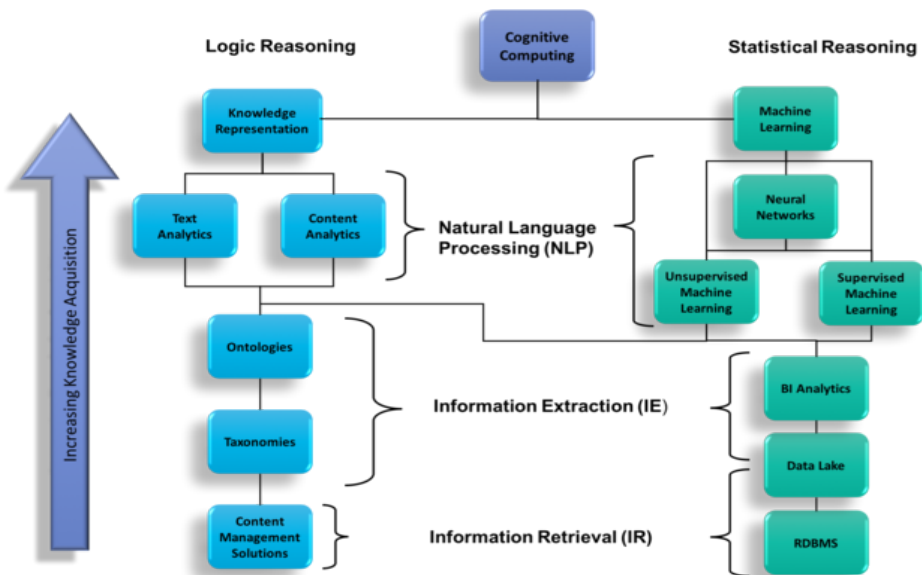


Si el supercomputador Watson, de IBM, es capaz de responder correctamente a la pregunta "cuáles son las mayores ciudades de Estados Unidos", es porque es capaz de encontrar muchas variantes de frases como "Nueva York es la mayor ciudad" o "no hay ciudad mayor que Nueva York" y porque dispone de un cuerpo (base de datos) muy grande de informaciones. Lo mismo hace Google cuando compara fotos y es capaz de distinguir entre un perro y un gato (y otros animales u objetos).

Las nuevas redes neuronales, que imitan la red cerebral sea en forma virtual (mediante programación en supercomputadores) sea en forma física en nuevos tipos de microchips, son las que permiten el mayor desarrollo y el mejor aprovechamiento de estos nuevos métodos. (Vea capítulo 4.)

2.5. Big data vs ontologías

Un problema profundo del aprendizaje de máquina y del *deep learning* (a la derecha en el gráfico que sigue) es que no es capaz de manejar con sentido común los desafíos racionales, como la lectura y comprensión de un titular de prensa. La "computación cognitiva", que se espera para el futuro, debe ser capaz de razonar, y para ello tiene que entender conceptos como el espacio, el tiempo, la masa y la inercia, para nombrar unos pocos. Esta es la razón por la cual el análisis de grandes datos (*big data*) tendrá dificultades para cumplir con la promesa de descubrimientos ricos, sobre todo cuando se trata de datos no estructurados.



Fuente: CIO.com

El análisis de grandes datos con el aprendizaje de máquina no utiliza relaciones conceptuales elaboradas, no se introduce en la semántica. Lo mejor que puede ofrecer es la capacidad de clasificar, sin ninguna capacidad de contextualización, que sería el valor real para los negocios. En cambio, las "ontologías" (estructuras de orden semántico) ofrecen la posibilidad de modelar relaciones conceptuales más elaboradas. Este enfoque se conoce como "modelado declarativo" y permite modelar problemas del mundo real. Se basa en la sintaxis: sujeto, predicado y objeto son la base de una buena estructura de oración³. El desarrollo de ontologías mediante la adopción de este modelo declarativo ha de recorrer aún un largo camino para proporcionar perspectivas contextuales al análisis de datos. Junto con las nuevas arquitecturas de redes neuronales quizás permitan por fin un real avance en una computación dotada de comprensión (MIT Technology Review, 2/04/2016).

Hace años que se conocen las ventajas de las ontologías y que se propone usarlas para desarrollar la "web semántica" pero parece poco realista esperar que todos los usuarios se plieguen a sus exigencias. Quizás podría la inteligencia artificial, en el futuro, operar la transición en forma automática.

2.6. Inseguridad

Quizás el mayor problema ligado a los big data, aparte de la pérdida de privacidad, es la escasa seguridad de estas bases de datos, como ha sido desvelado por una encuesta realizada por Unisphere Research y patrocinada por Oracle⁴. Solo una cuarta parte de los encuestados indicó que cifran todos los datos y solo un poco más de la mitad (56%) cifran al menos una parte de las copias de seguridad. 81% de los encuestados citan el error humano como mayor preocupación, seguido por 65% de posibles ataques internos, y 54% se preocupan por el abuso de los privilegios de acceso de su propio personal. 51% reconocen que no tienen garantías y 21% desconocen si tienen garantías para prevenir que un administrador o desarrollador pueda eliminar accidentalmente una tabla o que de forma inintencionada pueda provocar algún tipo de daño a las bases de datos críticas (Hispacec, 18/11/2014).

También se ha de tomar en cuenta que la internet pública de hoy no es confiable para ofrecer un nivel consistente de servicio, fiabilidad, seguridad y flexibilidad para la comunicación de máquina a máquina, como en la Internet de las Cosas. Se han propuesto aquí diversos protocolos pero no existe aún un estándar único, a pesar de que crece la demanda por una conectividad segura (DiarioTI, 1/12/2014).

La policía británica ha advertido que miles de enlaces a imágenes de videocámaras, cámaras de vigilancia e incluso monitores para bebés de decenas de países y territorios, entre ellos España, han aparecido en un sitio web administrado desde Rusia. Son 4.591 cámaras en EEUU, 2.059 en Francia, 1.576

en Holanda, 500 del Reino Unido, 378 de España. Y aparecen los enlaces clasificados por país y marca de las cámaras (El Mundo.es, 20/11/2014).

También se ha de tener en cuenta que los hackers tienden a abandonar los ataques individuales (como vía e-mails) prefiriendo atacar directamente los servidores de las empresas – especialmente del área financiera – para robar las bases de datos completas. Por ejemplo, en la “Operación High Roller” se vieron afectadas 60 entidades financieras de todo el mundo, víctimas de un ciberataque en el que se extrajeron 60 millones de euros (El Mundo, 21/06/2014). Kaspersky Lab también descubrió como más de 100 bancos en 30 países fueron afectados por un masivo robo electrónico a través de una sofisticada operación en varios pasos. Los *hackers* lograron penetrar en los computadores de empleados de bancos de Rusia, Suiza, Japón, Holanda y Estados Unidos, captar su modo de operar y luego imitar los patrones de sus transacciones para luego transferir dinero a cuentas falsas en Estados Unidos y China e incluso a cajeros automáticos en los que esperaban sus cómplices (The NewYork Times, 15/02/2015).

Los expedientes médicos también parecen ser material de interés: cerca de 80 millones de éstos, de ciudadanos de todo el mundo, fueron robados a una importante aseguradora en 2015 (Hipertextual, 5/02/2015).

2.7. Utilidad e inutilidad para investigadores

La BSA (*Business Software Association*), promotora del desarrollo de *software* a nivel mundial, encabeza una campaña de marketing para incentivar el uso del análisis *big data*, mostrando “sus potencialidades y en qué áreas ya ha comenzado a marcar diferencias”. “La innovación en *software* continua encendiendo avances sin precedentes que transforman el mundo a nuestro alrededor, empoderándonos como individuos y haciendo crecer nuestras economías”, declara. Y lo justifica en un informe denominado “¿Por qué son tan importantes los datos?”, disponible para descargar gratis desde su página web⁵, donde muestra cómo, mediante el seguimiento de más de 1.000 datos por segundo, se han logrado importantes avances en la medicina infantil, disminuir la congestión vehicular, aumentar el rendimiento de las cosechas desde Iowa a la India, desarrollar edificios de bajo consumo energético, hacer “ciudades inteligentes” y mejorar la planificación de las trayectorias de vuelo de los aviones... todos casos donde el análisis semántico parece muy poco relevante (FayerWayer, 11/11/2015).

Según el informe de OBS Business School de 2014, durante ese año el 73% de las organizaciones mundiales invirtieron en esta nueva área de negocio; especialmente los medios de comunicación y las tecnológicas (Hipertextual, 20/11/2015). Gartner también puso el aprendizaje de máquina en lo alto de los nuevos negocios y Lukas Biewald asegura que puede dar a cada empresa una ventaja competitiva pero que no es posible sin un adecuado entrenamiento (ComputerWorld, 20/10/2015). Se han multiplicado las *start-up* que ofrecen servicios de análisis de grandes datos (como QDAMiner, MetaMind, Dato y MonkeyLearn). Según Startupxplore, eran 93 los proyectos dedicados a la gestión de este tipo de datos en 2015 (Hipertextual, 20/11/2015).

Pero ¿el análisis de *big data* le puede servir a un investigador en comunicaciones?

2.7.1. Análisis de textos

Según los propulsores de las aplicaciones de analítica *big data*, el análisis de texto descubre en éstos el “qué”, el “quién”, el “dónde” y el “cuándo” de las conversaciones, lo cual requiere –en realidad– escasa capacidad semántica. Pero también permitiría extraer el “cómo” la gente se siente y “por qué” la conversación está sucediendo. Y esto permitiría que las conversaciones sean clasificadas y los temas de discusión identificados, lo que el programa de análisis automático debe lograr lo mejor posible, para así convertir los datos no estructurados en datos estructurados y permitir a las empresas “escuchar” globalmente las transmisiones que captan. Dice Jeff Catlin, CEO de Lexalytic⁶,

El análisis de texto puede ser utilizado para desarrollar una mejor comprensión de los gustos, disgustos y motivaciones del cliente. Cambiar los incentivos del programa de fidelidad para que coincidan con los deseos de los clientes puede mejorar las ventas y aumentar la fidelización de clientes (Inside Big Data, 5/06/2015).

Ha quedado demostrado, sin embargo, que los actuales sistemas de análisis automático de textos en *big data* están lejos de ser realmente útiles como métodos de análisis de contenido. Estos sistemas (salvo quizás el del supercomputador Watson de IBM) no tienen aún ninguna capacidad de crear categorías conceptuales a partir de lo leído. Solo pueden confeccionar estadísticas de frecuencia de las palabras y eventualmente de conexiones entre algunas en una misma oración (“segmentos”), con lo cual los

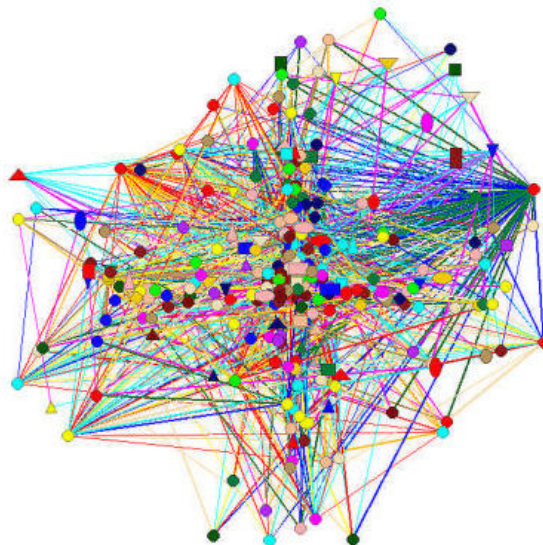
resultados entregados siguen siendo poco útiles y sin justificación alguna en el caso de los "segmentos" computados (siendo los algoritmos desconocidos y no revelados). Para obtener algo más útil, se requiere definir palabras clave, lo cual obliga al usuario a leer al menos parte de los textos... e introducir desviaciones. "No vamos a poder obtener ninguna información útil a no ser que codifiquemos de forma manual, al menos mínimamente, los textos a partir de un tesoro con su estructura jerárquica y correspondencias que, también, hemos de crear nosotros mismos. Y la pregunta es, ¿de qué nos sirve el software y Big Data entonces?" advierte F. López-Cantos en la Revista Latina de Comunicación Social (dic.2015). ¡Lo que se espera de la inteligencia artificial es justamente que la máquina nos evite este tedioso proceso!

Hice una posible excepción en el caso de Watson, el supercomputador de IBM que parece ser el más avanzado en materia de "comprensión" artificial. Es posible que Facebook, con sus actuales recursos de inteligencia de máquina, se le acerque. Es, al menos, lo que pretende, según las últimas noticias: anunció su sistema de análisis "Topic Data", que utiliza el análisis de texto para revelar lo que las audiencias están diciendo en su red sobre eventos, marcas, sujetos y actividades. Y lo ofrece a los vendedores, que pueden utilizarlo (arrendar el servicio) para "tomar mejores decisiones sobre sus productos y sus actividades" (Facebook, 10/03/2015). Pero sin duda, aquí, se definen inicialmente algunas variables o parámetros a considerar como marca, producto y criterios evaluativos.

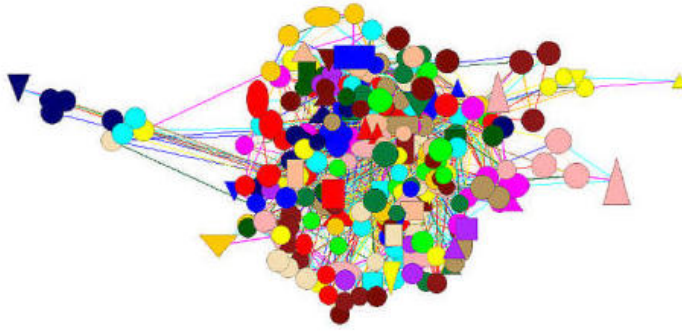
2.7.2. Análisis de bases de datos periodísticas

Lo que puede ser de particular interés para los medios de prensa es el análisis de sus bases de datos. Y en este caso puedo relatar una experiencia personal. En mayo del año 2000 empecé a recopilar y conservar en una base de datos de noticias relativas a los medios digitales, tomadas de más de 60 fuentes diferentes. Al llegar a 1.000 noticias, se hizo un primer análisis aplicando técnicas de minería de datos, experiencia de la cual se dió cuenta en el libro "Explotar la información noticiosa: Data Mining aplicado a la documentación periodística"⁷. En dicho libro, detallo ampliamente la metodología, las condiciones para el adecuado tratamiento de la información (conservando las noticias sobre la base de una estructura semántica), y el tipo de resultado esperable. La principal dificultad fue la escasa dimensión de la memoria RAM de nuestro PC para completar ahí todas las etapas, en particular el análisis multidimensional (en este caso 5 dimensiones, correspondientes a las 5 variables elegidas)⁸. Para poder mostrar, al menos en un par de gráficos, lo que se puede obtener con este análisis multidimensional, fue necesario trabajar con una muestra de solo 100 noticias. Se obtuvo una visualización interactiva, siendo posible la rotación en 3 dimensiones así como estiramientos que permitan poner en evidencia las confluencias y los casos marginales (Gráficos siguientes).

Cien noticias en cinco dimensiones

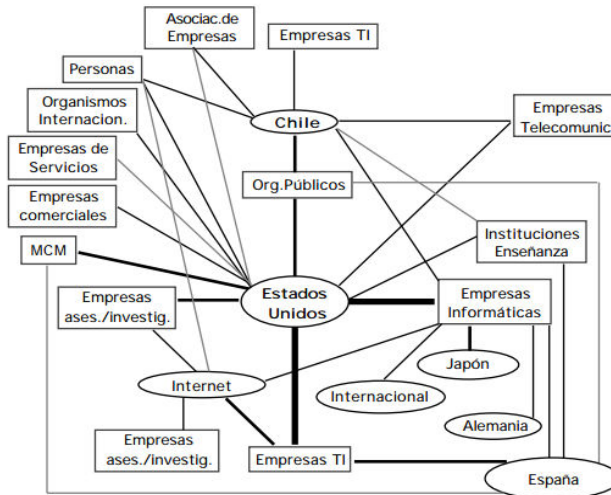


Visualización separando casos extremos

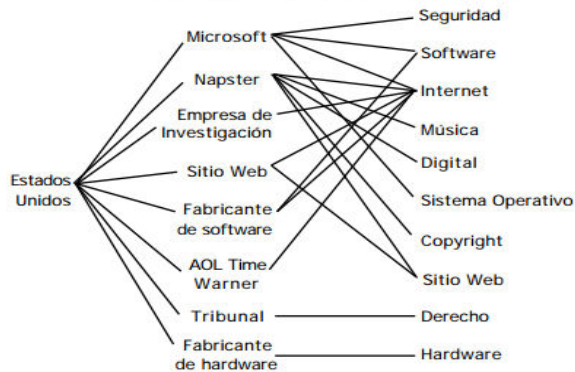


El análisis combinatorio de pares y tríadas de variables facilitó también la confección de gráficos a partir de las estadísticas obtenidas (ejemplos en gráficos siguientes).

Asociación entre Lugares y clases de Implicados



Tríadas Lugar-Implicados-Descriptorios



En noviembre de 2002 tenía ya 5.000 noticias estructuradas de un modo que facilitara el análisis estadístico y publiqué los resultados de éste en la revista Razón y Palabras⁹. En este caso extraje marginales simples (frecuencias) por categorías y sub-categorías así como frecuencias de combinaciones de dos valores de una misma variable – como pares de descriptorios temáticos (coocurrencias) – y de dos variables, como lugares o “implicados” (empresas) combinados con descriptorios temáticos, lugares y fechas, etc.

Aquí algunas de las conclusiones de estas experiencias (extraídas del libro):

Los resultados expuestos son comúnmente imposibles de obtener en sistemas de documentación periodística concebidos de otra manera¹⁰ y sin las herramientas aquí descritas. (p.186)

Sin duda los primeros hallazgos - que proceden de la confección de las estadísticas iniciales - no presentan grandes sorpresas, pero al avanzar en los procedimientos y en la complejidad del análisis hemos visto aparecer otras informaciones que aportaron el conocimiento nuevo que estábamos buscando. (p.187)

El gráfico pone mucho mejor de manifiesto los múltiples vínculos que asocian los valores de los diversos atributos, o sea de la manera en que las noticias pueden relacionarse entre sí. La visualización, no sólo es útil para determinar áreas que requieren mayor profundización del análisis o ajustes en la secuencia y las etapas de reiteración propias de este tipo de estudio: permite además descubrir tanto agrupaciones muy peculiares (los *clusters*) como casos excepcionales, lo que – en ambos casos - constituye un metaconocimiento de la mayor importancia. Ésta es una importante ventaja.

Sin embargo, una gran cantidad de datos produce rápidamente una saturación del espacio visual disponible (habitualmente una pantalla de ordenador) y es indispensable recurrir a operaciones lógicas complementarias, tales como la fijación de umbrales y la eliminación de situaciones intermedias, para encontrar y exhibir las agrupaciones más fuertes o bien los casos marginales que pudieran ser indicadores de casos extraordinarios particularmente relevantes o indicadores de errores de procedimiento en la recopilación de la información que ingresa a la base de datos. Si bien estos procedimientos restrictivos forman parte de la metodología y llevan a obtener los nuevos conocimientos buscados, tienen generalmente el costo de ser muy difícilmente graficables en presentaciones bidimensionales.

Es una de las razones por las cuales las “suites” de *Data Mining* incorporan facilidades asociadas a la gráfica tridimensional, como la rotación en pantalla, y - las más avanzadas - incluso técnicas de realidad virtual, que permiten manipular directamente los componentes mediante el guante digital, visualizar el cubo mediante visión estereoscópica (con un visor especial) e incluso “penetrar” en él para mirarlo desde diversos ángulos desde su interior y no solamente desde el exterior. (pp.190-191)

3. Algoritmos

3.1. La invasión de los algoritmos

Los algoritmos son las secuencias de órdenes que conforman la base de la programación de los computadores y, ahora, de las respuestas que obtenemos cuando navegamos por internet.

Tendencias · Cambiar

#PequeñosPlaceresDeLaVida
13,1 K Tweets

#BienvenidoAbril
40,9 K Tweets

#SueldoEtico400mil
Comenzó a ser tendencia en la última hora

#PremiosTVyNovelas
31 K Tweets

#JaimeGuzman
Siendo tendencia desde hace 6 horas

Rafael Dudamel
16,3 K Tweets

HOY JUGAMOS A TU LADO VIGNA
21,8 K Tweets

Las redes sociales utilizan los algoritmos para ordenar lo que nos muestran al abrir su página. Así, Twitter cambió la forma de presentación cronológica de los tuits por una basada en la estadística ("los más vistos"), es decir basados en un algoritmo (aunque se puede desactivar). Instagram anunció que pronto haría lo mismo pero, de momento, se retractó al recibir numerosos reclamos. Facebook selecciona los "News Feeds" del mismo modo desde hace tiempo y lo hace ahora con las actualizaciones de estado (Fortune, 28/03/2016).

Las "tendencias" que Twitter parece extraer "objetivamente" a partir de los retuits (y otras señales), son una buena muestra de mecanismo distorsionador: una vez que un tema adquiere este codiciado status, atrae aun más atención y se produce un efecto "bola de nieve" que lo viraliza (dentro y fuera de Twitter). El riesgo es que los periodistas que utilizan Twitter como fuente sean poco precavidos y caigan en la trampa de ver ahí un "tema importante".



Algo más elaborado ha sido el proyecto Tay de Microsoft (ver al lado, Twitter-Profil-der-Microsoft, fuente: proyecto Tay). Basado en inteligencia artificial, pretendía entablar una conversación "casual y fluida" con jóvenes entre 18 y 24 años. Pero se acumularon los mensajes ofensivos enviados por éstos y, en menos de 24 horas, pasó de decir que los humanos eran "súper guay" a volverse cada vez más xenófobo, malhablado y sexista. Se hizo "simpatizante de Hitler y acabó deseando a muchos que acabasen en un campo de concentración" dice la nota de prensa (El Mundo.es, 28/03/2016). Así, Microsoft tuvo que desactivar la IA y pedir disculpas. La noticia acerca del racismo de su inteligencia

artificial ha dado la vuelta al mundo (y puso de cabeza a Microsoft). En diciembre 2016, sin embargo, Microsoft volvió a la carga con "Zo", un nuevo chatbot basado en IA, pero que limita fuertemente los temas que pueden ser abordados: nada de política... ni hablar de Tay (FayerWayer, 6/12/2016).

Kate Crawford¹¹, investigadora en el campo del aprendizaje automático, advirtió del peligro de los errores de los algoritmos en una reflexión en la que calificaba a este descubrimiento como la punta de un iceberg gigante, cuyo impacto puede ser brutal (Javier Pastor, Xataka, 23/11/2016). Crawford tomó como muestra el error de Facebook, que censuró la célebre foto titulada "The terror of war", de Nick Ut, tapando a la niña desnuda quemada con napalm, considerando que esta era una violación de la tradicional política contra los desnudos de niños. Esto llevó a la propia primera ministra de Noruega, Erna Solberg, a sublevarse, indicando que, de este modo, Facebook estaba – ilegítimamente – "redactando nuestra historia compartida". Los errores de Microsoft, Facebook y Google (que exhibió fotos de mujeres afroamericanas en respuesta a una búsqueda de fotos de gorilas) son sin duda casos extremos, pero también una muy buena demostración de lo que puede pasar con cualquier algoritmo, desde el de Google Search al de Facebook y todas las aplicaciones de análisis de "big data" que tanto se "marketean" actualmente.



3.2. Como distorsionan el conocimiento

Los algoritmos no son en sí-mismos los responsables de generar ocasionalmente errores graves: los responsables son quienes los programan.

Los algoritmos no existen en el vacío. Requieren seres humanos. ...

La verdad es que muchos de los algoritmos informáticos de hoy en día son imprecisos, aunque las empresas y los codificadores están siempre trabajando para mejorarlos. Y porque son imprecisos, necesitan la ayuda de seres humanos. ...

Incluso si había un sesgo anti-conservador [en Facebook¹²], este es el tipo de cosa que sucede en cualquier servicio web, ya sea Google o Amazon o The New York Times o Wired. Eso es porque los seres humanos son parciales. Y eso significa que las empresas también tienen un sesgo. ...

[Facebook, Google y la prensa tecnológica] pasaron años alimentando la idea de que Internet es totalmente automático. A pesar de que no funciona de esa manera, la gente quiere que lo haga. Cuando alguien sugiere que lo hace, la gente tiende a creer que lo hace. (Wired, 2/09/2016)

El efecto distorsionador de los algoritmos es fácil de observar haciendo una búsqueda en Google, que selecciona los resultados de nuestras búsquedas utilizando un algoritmo ultra-complejo (más de 200 reglas). La empresa, "para facilitar las búsquedas" ha instalado un sistema de predicción de los términos de búsqueda, basado en un algoritmo que toma en cuenta lo más buscado pero también muchos otros factores (como el país del lector, la lengua, sus intereses registrados, y al menos otros doscientos factores), en algo que se combina además con sus propios sistemas de clasificación. (Ejemplo a continuación).



Fuente: búsqueda propia

Dice Morozov:

Lo que resulta de veras irritante es que Google insista en la supuesta neutralidad y objetividad de sus algoritmos. En lugar de reconocer que estos pueden tener dificultades y sesgos que es necesario corregir, Google se comporta como si introducir a los humanos para que revisen de tanto en tanto el trabajo de sus algoritmos equivaliera a abandonar toda fe en la inteligencia artificial como tal.

Resistir a reconocer que sus algoritmos en ocasiones pueden tener un mal funcionamiento le permite a Google zafarse de una serie de aspectos éticos un poco complejos sobre su trabajo. (E.Morozov, *"La locura del solucionismo tecnológico"*:167)

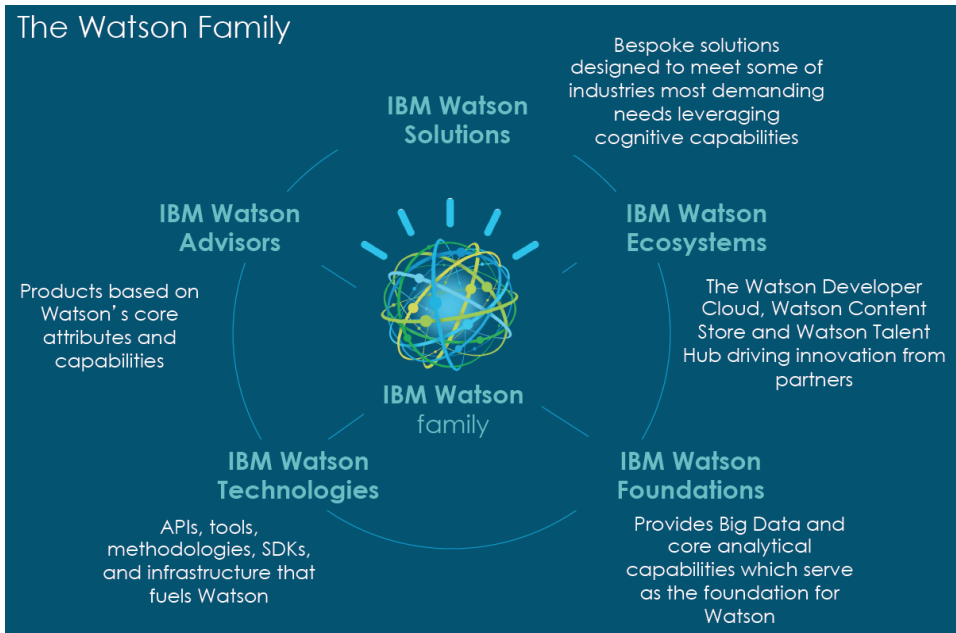
El problema ético al que alude Morozov tiene que ver con la relativa facilidad para torcer los resultados de búsqueda, por ejemplo utilizando "generadores de búsquedas" que repiten centenares de veces términos que pueden ser ofensivos:

Supongamos que un enemigo suyo, en un esfuerzo deliberado por manchar su reputación, decide pagarles a los usuarios para que busquen su nombre seguido de la palabra «pedófilo». Un ejército de entusiastas colaboradores, reclutados a través de sitios como Craigslist y Mechanical Turk, de Amazon, está generando el volumen suficiente de búsquedas para que esa palabra reemplace otros términos más positivos que se han asociado a su nombre. Ahora, cualquiera que lo busque también sabrá que quizá usted sea un pedófilo y, recuerde, no hay manera de apelar porque los algoritmos de Google son los que están a cargo y jamás se equivocan (Morozov: 168).

Es inútil tratar de reclamar, porque Google (y otros inventores de algoritmos) pretenden que "reflejan objetivamente" lo que está en la web. Como agrega Morozov, *"la compañía no solo refleja, además da forma, crea y distorsiona, y lo hace de múltiples maneras, que no pueden reducirse a una única lógica de internet"* (p.170). ¡Internet, como tal, no es una "fuerza social" como parecen creer! Los motores de búsqueda y los algoritmos crean una nueva realidad.

Google debería dejar de esconderse detrás de la retórica de los espejos y los reflejos, reconocer su enorme papel en la configuración de la esfera pública y comenzar a ejercerlo con mayor responsabilidad. Ser «objetivo» es una tarea muy difícil; no sucede con naturalidad después de haber delegado todo el trabajo a los algoritmos. Los nuevos jefes supremos de los algoritmos no deberían aspirar a ser autómatas éticos; solo siendo autorreflexivos y moralmente imaginativos pueden estar a la altura del enorme peso de sus responsabilidades cívicas (Morozov: 171).

Lo más grave es que se altera de este modo el conocimiento a nivel global. Y podríamos ir de mal en peor, porque hoy en día los algoritmos son la base del aprendizaje automático que está más que nunca al alcance de cualquier programador y que es también la base del análisis de "big data". Para experimentar con estos servicios tenemos plataformas como IBM Watson Developer Cloud (ver ilustración), Amazon Machine Learning, Azure Machine Learning y TensorFlow, y todos hacen ingentes esfuerzos de mercadeo en la red para convencer a todas las empresas, hasta las más pequeñas, de las "bondades" y oportunidades que su uso ofrecería.



Fuente: *Softwareinsider.org*, 2014-01-10

Imagine ahora que los medios de prensa empiecen a utilizar este tipo de análisis, basándose en lo que más leen sus lectores: orientarán su selección y sus reportajes en función de lo preferido por la mayoría. Y, como se sabe que las noticias deprimentes gustan poco y no se reenvían (recuerde que los ponen también en Facebook y Twitter), terminaremos con un panorama de "mundo feliz" irreal.

3.3. ¿Algoritmos no supervisados?

Los algoritmos analíticos son la base de una de las líneas de desarrollo de la inteligencia artificial, el llamado "aprendizaje supervisado" porque –al menos– los humanos deben intervenir para dar inicialmente a la máquina algunos modelos de lo que se podría considerar útil. El principio básico, tomado del modelo animal (puesto en evidencia por Pavlov), es el aprendizaje por refuerzo, es decir por medio de la repetición, lo cual es fácil de traducir en un mecanismo estadístico, pero es claramente más mecánico que inteligente.

Para superarlo se está avanzando hacia un "aprendizaje no-supervisado", donde se pretende que la máquina saque conclusiones "sin intervención humana" (¿pero quien la programa sino un humano?). Y "ya existen compañías que se centran completamente en enfoques de aprendizaje automático no supervisado, como Loop AI Labs, cuya plataforma cognitiva es capaz de procesar millones de documentos no estructurados y construir de forma autónoma representaciones estructuradas" (Xataka, 28/03/2016). (Ver ilustración siguiente). Se dice que operaría como el cerebro humano y que la máquina "comprendería" el contenido semántico y hasta "los motivos de una persona". ¿En serio?

LOOP COGNITIVE COMPUTING PLATFORM

The flagship Loop AI Labs product is the [Loop Cognitive Computing Appliance](#), a fully unsupervised next-generation machine intelligence technology, consisting of proprietary algorithms and low-cost, custom-built High Performance Computing GPU accelerated hardware.

It allows organizations to capitalize on the power and potential of their [dark data](#) and it automates processing and understanding of the dark data specific to each enterprise, breaking prevailing barriers that previously required tradeoffs among speed, cost, and quality.

The [Loop Cognitive Computing Appliance](#) works on principles inspired by the neocortex, the brain's center for language and reasoning. Loop AI's homegrown algorithms learn the underlying

Fuente: Amazonaws.com

¿Que podemos esperar si, como se lo proponen los expertos en inteligencia artificial, se pasa de este aprendizaje de máquina supervisado a uno no-supervisado en que la máquina "decidiría" lo que es importante? ¿En base a qué criterios? ¿Funcionaría sin algoritmos? Se habla de "algoritmos capaces de aprender": pero son algoritmos capaces de generar otros algoritmos. Y un programador debe haber determinado el qué y el cómo. Ningún computador puede crear su propia programación. Incluso las redes neuronales (que pretenden emular el cerebro humano) deben ser diseñadas, incluir reglas de valoración y luego entrenadas, al menos para verificar que operen adecuadamente (¿y quien determina lo "adecuado"?).

Los algoritmos de análisis de grandes datos, además de incluir reglas de selección (que deberían ser explícitas pero permanecen generalmente ocultas) deben recurrir a la estadística. ¿Pero que estadística? ¿Solo de frecuencias? ¿De correlaciones (paramétrica) o de relaciones asociativas (no-paramétrica)? Se ofrece mayoritariamente a las empresas para ayudar en la toma de decisiones, e.d. con una función predictiva, lo cual transforma el análisis en causante de decisión. ¿Es esto correcto? (¡Es dudoso que lo sea!)

La mente tiende automáticamente a deducir causalidad donde solo hay algún tipo de correlación. Además, las reglas de la estadística dicen que los resultados obtenidos de muestras grandes son más confiables que de muestras pequeñas, lo que es la base del marketing de esta metodología pero aumenta los riesgos de mala interpretación de los resultados del análisis (por exceso de confianza). ¿Son sensibles los científicos de datos a estas trampas? "*La tendencia a ver patrones en lo aleatorio es abrumadora*", nos advierte Daniel Kahneman (p.157), tanto como de ver causalidad donde no la hay. "*Nuestra predilección por el pensamiento causal nos expone a serios errores en la evaluación de la aleatoriedad de sucesos realmente aleatorios.*" (*ibidem*:155)

Tim O'Reilly propuso un conjunto de cuatro reglas para lograr saber si podíamos o no fiarnos de un algoritmo:

1. Sus creadores han aclarado el resultado que buscan, y es posible que observadores externos verifiquen ese resultado.
2. El éxito es medible.
3. Los objetivos de los creadores del algoritmo están alineados con los objetivos de los consumidores del algoritmo.
4. ¿Permite el algoritmo llevar a sus creadores y usuarios a una mejor toma de decisiones a largo plazo? ("*The great question of the 21st century: Whose black box do you trust?*", O'Reilly.com, 15/09/2016)

Aunque se diga que “No tenemos más remedio que recurrir a máquinas y además necesitamos que estas máquinas sean capaces de interpretar los datos, comprenderlos y sacar conclusiones de forma inteligente” porque “existe un gran valor oculto en los millones y millones de toneladas de datos que una organización puede recabar” (Xataka, 28/03/2016), es el ser humano que aporta la inteligencia, debiendo evaluar estas conclusiones. Morozov nos aconseja desconfiar:

Debemos dejar de creer que los nuevos filtros y prácticas algorítmicas impulsados por los recientes intermediarios digitales (y sus entusiastas genios de internet) son apromblemáticos, objetivos y superiores por naturaleza a los filtros y prácticas que los precedieron. Esos nuevos filtros podrán ser más veloces, baratos y eficaces pero la velocidad, el costo y la eficacia guardan solo una relación periférica con las funciones civiles que los nuevos filtros y algoritmos desempeñarán en nuestra vida. Si no sometemos a esos filtros más veloces, baratos y eficaces al estricto escrutinio ético que merecen, corremos el riesgo de caer en una de las tantas falacias del solucionismo y celebrar avances relacionados con problemas menos importantes, al tiempo que desatendemos cuestiones más acuciantes aunque menos obvias (Morozov: 174).

La cultura misma puede verse afectada en forma global, sobretodo si se deja que sean los algoritmos que determinen lo “más interesante” sobre la base de lo más consultado, leído, visto u oído. Los “clicks” y “me gusta” tienden a concentrarse en opiniones no calificadas mientras las críticas de expertos resultan poco atractivas para la mayoría. Agrega:

Aquí la lógica de funcionamiento es simple: la época previa a internet se asocia al conocimiento experto, y la época posterior a internet se asocia al populismo; nosotros estamos en la época posterior, ergo: debemos ser populistas. [...] De acuerdo con esta lógica – que elogia la participación cultural masiva como algo bueno en sí mismo, sin importar sus efectos en la cultura – incluso las calificaciones de discos y canciones que generamos en iTunes y Spotify podrían terminar siendo elegidas en lugar de las que producen los críticos profesionales. [...] La socióloga Nancy Hanrahan es especialmente perceptiva en este aspecto:

“No se puede negar que la erosión del conocimiento cultural experto posibilitada por las nuevas tecnologías es democrática, al menos, en un sentido: hay una participación más amplia en la creación y evaluación de la cultura. Si fuera tan fácil cuantificar la democracia como lo es cuantificar las ventas de discos o la cantidad de clics que recibe un sitio web, quizás la discusión terminaría allí. Pero no es solo la participación sino también los términos de la participación los que deben tenerse en cuenta. Si una mayor participación en la cultura a través de las tecnologías digitales y las estructuras de red en las que se hallan insertas favorece al mercado, desalienta la innovación artística o puede comprarse a costa de la reflexión crítica sobre el arte, ¿con qué argumentos puede considerarse democrática? Si, por otro lado, la democracia significa la ampliación de oportunidades de deliberación, de vida pública o de genuina diversidad, la situación actual se queda a mitad de camino.”

Otro punto clave que suelen pasar por alto los solucionistas es que los críticos de música profesionales – y esto también vale para los críticos de cine y los literarios – cumplen muchas otras funciones que no pueden delegarse a las masas con tanta facilidad. Una de ellas es identificar creaciones musicales innovadoras, incluso provocadoras, y de alta calidad, y explicarle al público por qué son merecedoras de su apreciación (Morozov, pp.204-205).

3.4. Personalización de algoritmos

Se anunció que 2016 se convertiría en el año de los bots inteligentes (software con inteligencia artificial que se puede programar para hacer todo tipo de cosas) (Capital.cl, 5/04/2016). Es decir que los algoritmos, especialmente los de aprendizaje de máquina, se multiplicarían. Satya Nadella, CEO de Microsoft, es uno de sus promotores y anunció que su compañía, a pesar del fracaso de Tay, se propone ofrecer lo que denomina el “Microsoft Bot Framework”: un conjunto de herramientas de código y programas de aprendizaje de máquina para que cada cual pueda construir su propio sistema, desde motores de búsqueda hasta asistentes personales que respondan al nombre del usuario y sus intereses (The Verge, 30/03/2016).

Se podría considerar que los algoritmos de recopilación y análisis de intereses personales –individuo por individuo– (los “little data”) podrían ser una real ayuda para los usuarios y es posible que, efectivamente, puedan facilitar servicios personalizados. Pero sigue siendo dudoso que sean legítimos si se imponen sin consultar al usuario como ocurre actualmente en la gran mayoría de los casos.

Siempre he sido partidario de la personalización en el caso de los periódicos digitales pero también he recalcado la importancia de seguir incluyendo información sobre otros temas, especialmente los que pueden ser relevantes para toda la población, para asegurar una adecuada participación ciudadana. Establecer las prioridades en este campo más general ha sido siempre y deberá seguir siendo el rol de los equipos editoriales de cada medio.

En otras palabras, el tema de la personalización no puede ser tratado al margen del tema de la vida social, del bien común. Se trata, en esencia, de un tema de ética de las comunicaciones, algo que – desgraciadamente – parece ser ignorado por la concepción actual del mercadeo digital, influenciado por la fe ciega en la "objetividad" de los algoritmos. Y si esto ya es éticamente cuestionable, no olvidemos lo fácil – y perverso – que es introducir un pequeño factor que favorezca (invisiblemente) los intereses de una empresa o una facción política.

Podemos creer que estamos en una nueva "era del conocimiento" y es obvio que tenemos más que nunca acceso a innumerables fuentes de información. ¿Pero tenemos realmente acceso a lo más significativo y lo más beneficioso (para nosotros y para la comunidad)? ¿O solo a lo que algunos parecen preferir y a lo que los creadores de algoritmos creen más "adecuado" o más "representativo" según sus propios (y desconocidos) criterios?

Se dice – sin duda correctamente – que permitir a las empresas el acceso a los datos personales es la forma adecuada de remunerar los servicios que presentan como "gratuitos". ¿Pero no sería más correcto que cada uno pudiera decidir qué datos liberar y qué intereses priorizar en vez de obligarnos a "abrir completamente la puerta" para que una empresa analice todo nuestro comportamiento en línea (incluidos los contenidos de nuestros mensajes personales, e.d. violando sistemáticamente el secreto de la correspondencia)? Y si rehusamos, actualmente, nos condenamos al ostracismo. ¿No es ésta una nueva forma de esclavitud?

Debemos considerar también que compañías como Facebook han comenzado a utilizar algoritmos para predecir qué usuarios podrían cometer delitos haciendo uso de sus servicios. En 2012, Reuters informó que Facebook, provisto de sus algoritmos predictivos, descubrió a un hombre de edad media que mantenía conversaciones de contenido sexual con una niña de trece años y hacía planes para encontrarse con ella al día siguiente. Avisó a la policía y ésta lo detuvo. Es difícil cuestionar la aplicación de tales métodos para la aprehensión de este tipo de delinquentes. Pero también podría detectar traficantes de drogas y hasta posibles infractores de derechos de autor. ¿Las redes sociales convertidas en policías?

No sabemos si Facebook tiene algún tipo de Pedofilómetro, pero teniendo en cuenta el amplio análisis de usuarios que ya realiza, no le resultaría muy difícil crear uno, y no solo para hacer un ranking de pedófilos. ¿Qué tal un Drogómetro? ¿Y un Comunistómetro?, a Joseph McCarthy encantaría la idea. Con los datos suficientes y los algoritmos correctos, todos corremos el riesgo de ser sospechosos. ¿Qué sucederá cuando Facebook nos entregue a la policía sin que hayamos cometido ningún delito? (Morozov:216)

Se ha sugerido la creación de agencias especializadas e independientes encargadas de auditar los algoritmos. La idea es sin duda interesante pero plantea varias interrogantes, no siendo la menor quienes las financiarían.

3.5. La pesadilla de la identidad alterada en la red

Las empresas presentes en internet recopilan, conservan y analizan cada vez más datos acerca de quienes visitan sus páginas o sus inserciones en redes sociales. Todo para "mejorar" su acceso y sus ofrecimientos a sus posibles clientes. De ahí que el análisis de grandes datos (*big data*) forme parte del trabajo de los departamentos de mercadeo y que haya una demanda creciente (insatisfecha) por "científicos de datos". Sin embargo, los que ya trabajan en esta área se están dando cuenta de que no sólo es complicado el análisis sino muy difícil obtener resultados claramente útiles porque los usuarios de internet tienden a falsificar sus datos personales.

Acaban de darse cuenta que la "personalidad digital" (o virtual) puede ser muy diferente de la real. Nuestro perfil y nuestros posts en nuestros "muros" conforman los datos que, libremente, publicamos y que pasan a engrosar las bases de datos de las redes sociales en las que participamos y de las empresas que nos observan. Éstas pueden conocer nuestros gustos, pero solo en función de lo que exhibimos directa o indirectamente. Nuestras búsquedas y nuestras compras, por ejemplo, pueden dar una idea

de lo que nos interesa y ésto puede pasar a formar parte de nuestra identidad virtual aunque no signifique que realmente nos sea propio (podemos comprar regalos, consultar cosas que interesan a nuestros amigos, etc.). Y podemos, obviamente, declarar gustos e intereses falsos y esconder los verdaderos. La gestión de la reputación, en tanto selección de la información personal que se expone públicamente, es un elemento importante en la manera en que la gente funciona en las redes, establece sus credenciales y construye confianza. Y lo propio de las redes sociales es que los individuos interconectados pueden crear "identidades digitales" verdaderas o bien parcial o totalmente ficticias, como deseen (Rainie y Wellman, pp.15-17).

Estas falsedades o "datos sucios" ("*dirty data*") constituyen una parte cada vez más problemática de los *big data*.

Los consumidores están llenando las bases de datos de las compañías con una avalancha de datos falsos, de medias verdades y de datos que se ajustan más a lo que la empresa puede querer oír que a lo que el consumidor querría decir.

Más de la mitad de los consumidores están, de hecho, dando ya datos y llenando las bases de datos de las compañías de mentiras cuando se aborda la información que estos dan a las empresas. De forma más exacta y concreta, y según un estudio de Verve, el 60% de los consumidores provee de información falsa a las empresas que le piden sus datos personales en internet. El estudio está realizado sobre una muestra británica, pero sirve para apuntalar la percepción de que existe un problema en el mundo de los datos.

Un 81% de los encuestados pone su privacidad como una de sus principales preocupaciones, un 77% señala que las peticiones de información son intrusivas y un 76% innecesarias. (PuroMarketing, 2015)

Un estudio de Shop+ determinó que los "*millennials*"¹³ mienten –como muchos otros– sobre su fecha de nacimiento y, especialmente, cuando les permite beneficiarse de un servicio extra o una oferta (*ibidem*). Dar un e-mail falso es también común, para evitar el ser perseguido por anuncios y spam.

Ésto podría ser preocupante para el futuro, especialmente si no se toman mejores medidas para resguardar la privacidad de los datos, como nos advierte Eric Schmidt:

En el futuro la identidad será la materia prima más valiosa para los ciudadanos y ésta existirá principalmente *online*. La experiencia *online* comenzará al nacer, o incluso antes.

En respuesta, las compañías tendrán que crear nuevas herramientas para controlar la información, tales como listas que permitan a la gente determinar quien ve sus datos. Las tecnologías de comunicación que utilizamos hoy en día son invasivas por diseño, recopilando nuestras fotos, comentarios y amigos y almacenándolas en bases de datos gigantes en las que se puede buscar, y en ausencia de una regulación exterior sólo resta confiar en la buena fe de empleados, personal burocrático de las universidades y otras personas. (Schmidt y Cohen:58)

Pero no todos tienen buena fe y no faltan quienes no solo modifican sus propios datos sino que buscan una forma de desaparecer y hacerse de otra personalidad:

La identidad *online* se convertirá en una moneda muy potente que incluso verá cómo se crea un nuevo mercado negro donde la gente puede comprar identidades reales o inventadas. Los ciudadanos y delincuentes se verán igualmente atraídos por una red así, donde la identidad falsa podría proporcionar cobertura para un traficante de drogas, así como refugio para un disidente político. La identidad será fabricada o robada y se ofrecerá completa con datos de fechas anteriores, registros de actividad IP (Protocolo Internet), amigos falsos y compras de artículos, así como otros medios para hacer aparecer convincente la información. (*ibidem*:61)

¿Quién tomará "el toro por las astas" e introducirá los necesarios mecanismos de verificación al mismo tiempo que reglas mucho más estrictas de resguardo de los datos personales? Por ahora, salvo algunas reglas básicas establecidas en la Unión Europea, nada está claro al respecto. Aún así, algunos –como el mismo Eric Schmidt– son optimistas en cuanto a la detección de mentiras e identidades falsas: hay tanta información en internet y habrá tantos medios de reconocimiento facial en organismos públicos que será cada vez más difícil esconder la verdad (Nos hacen pensar en la serie de televisión "*Person of Interest*").

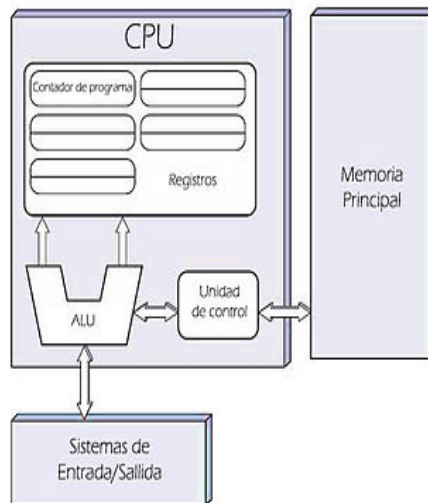
4. Inteligencia artificial

4.1. Arquitecturas de aprendizaje profundo

El aprendizaje profundo sólo se convirtió en relevante en los últimos cinco años, después de que los investigadores se dieran cuenta de que los chips originalmente diseñados para manejar gráficos de videojuegos eran más potentes y podían servir para este tipo de procesamiento. Con ellos es posible simular las redes cerebrales, lo cual ha permitido avances rápidos en el aprendizaje de máquina y el renacimiento del *software* informático que pretende resolver los problemas un poco como un cerebro humano, mediante un complejo proceso de reconocimiento de patrones distribuidos a través de muchos nodos o "neuronas" virtuales. El actual poder de cómputo ha permitido a las redes neuronales simuladas reconocer imágenes, palabras y caras, así como pilotar coches de auto-conducción y ganar en Go y Jeopardy. Han sido progresos notables. Pero aún falta mucho para una real imitación del cerebro humano. Dos son los caminos abiertos hoy.

4.1.1. Lo tradicional: paralelismo masivo

Los más potentes supercomputadores se construyen de acuerdo a la arquitectura tradicional, establecida por von Neumann en 1957 y que encontramos en nuestros PC y *smartphones*. Pero la gran diferencia es que incluyen una gran cantidad de procesadores que funcionan en forma paralela. El "paralelismo masivo" es lo que les da su poder superior, al permitir numerosas operaciones simultáneas. Agregándoles una memoria extremadamente amplia, se logran avances en IA como los de Watson, el supercomputador de IBM, que puede procesar 500 gigabytes por segundo (el equivalente de un millón de libros).



Watson lleva "entrenándose" desde 2007, acumulando conocimientos culturales y la jerga propia de cada uno de los sectores en los que se va a introducir, ya sea la sanidad, las finanzas o la gastronomía.

En palabras de Raúl Fernández, uno de los investigadores de IBM, Watson ya maneja el lenguaje «como lo haría un niño precoz». Y, como ya ha desvelado la CEO de la compañía, Ginni Rometty, Watson podría ya formar parte de nuestra actividad diaria sin que lo supiéramos. Además de por grandes empresas como Hilton o la cadena de almacenes Macy's, sus capacidades han sido incorporadas ya por sistemas como los que ayudan a los médicos a realizar diagnósticos, al robot de banca Pepper, que la compañía Soft-Bank Robotics desarrolla para mejorar la relación con los clientes del sector, y a la aplicación meteorológica de The Weather Company, recientemente adquirida por IBM. (PC World, 11/11/2016)

Según IBM, usa más de 100 técnicas diferentes para analizar el lenguaje natural, identificar fuentes, encontrar y generar hipótesis, buscar y puntuar evidencias, combinar y clasificar hipótesis.

4.1.2. La simulación de redes neuronales

El aprendizaje profundo (*deep learning*), el mayor avance actual en inteligencia artificial, implica que los datos pasan a través de grandes conjuntos de neuronas simuladas en grandes computadores de arquitectura clásica. Avanzar aquí supone contar con procesadores mucho más potentes.

La compañía Nvidia, cuyo *hardware* (procesadores gráficos) sustenta gran parte del progreso de los últimos años, ha creado un chip para mantener en marcha el progreso, el P100. Nvidia anunció en 2016 un nuevo chip llamado "Tesla P100", que está diseñado especialmente para disponer de más poder para el aprendizaje profundo. Nvidia gastó más de 2 millones de dólares en I+D para producir este nuevo chip, que cuenta con un total de 15 mil millones de transistores, aproximadamente tres veces más que sus chips anteriores. Una red neuronal artificial impulsada por el nuevo chip podría aprender de los datos de entrada 12 veces más rápido que los anteriores. El P100 ya está en producción y se espera que las empresas de computación en la nube puedan empezar a usarlo este mismo año. Permitiría a los informáticos alimentar con más datos a sus redes neuronales artificiales o crear mayores conjuntos de neuronas virtuales (MIT Technology Review, 6/04/2016).

4.1.3. Las neuronas artificiales

La mayoría de los científicos de computación piensan que es sólo el comienzo de lo que será posible. Este tipo de *hardware* se ajusta muy poco en su arquitectura a lo que son las redes neuronales biológicas. Esto significa que puede tomar días o incluso semanas entrenar una red neuronal para resolver un solo problema incluso en un clúster de computadores y luego requerir una gran cantidad de energía para resolver el problema una vez que están entrenados.

La solución la quieren abordar en IBM, volviendo a las descripciones originales de las redes neuronales artificiales que se deben a Frank Rosenblatt, psicólogo de Cornell, que creó en 1959 el "perceptrón", una red de 400 células fotoeléctricas conectadas, destinada a reconocer patrones visuales sencillos (letras). Lamentablemente, Minsky y Papert dedicaron luego un libro a demostrar la incapacidad de los perceptrones para realizar las tareas más sencillas y el interés decayó. Resurgió a fines de los años setenta y ochenta pero se orientó hacia las redes simuladas. Los investigadores de IBM quieren cambiar ésto perfeccionando otra tecnología. Se llama computación resistiva y consiste en tener pequeñas unidades de cómputo de naturaleza analógica, que pueden retener su historia para que puedan aprender durante el proceso de entrenamiento.

Acelerar las redes neuronales con el *hardware* no es nuevo para IBM: ya tiene los chips neuromórficos TrueNorth pero solo útiles para resolver problemas de diseño de redes neuronales. En éstos, el entrenamiento y la "ponderación" de los parámetros necesarios en las conexiones (synapsis) aún deben hacerse en computadores normales.

Los nuevos chips "resistivos" (*Resistive Processing Units*, RPUs) imitan una sinapsis cerebral: reciben una multiplicidad de entradas analógicas (voltajes) y utilizan una función de "ponderación" basada en la experiencia pasada para determinar qué debe ser traspasado a la etapa siguiente. Como no necesitan una conversión analógica-digital ni consultar una memoria externa, son sumamente rápidos y requieren poca energía. El trabajo reciente ha permitido su operación en una rejilla de 4.096 x 4.096 elementos. Los RPUs siempre tienen cierto grado de imprecisión pero se han encontrado métodos para obtener resultados con un grado de imprecisión aceptable.

Esta nueva arquitectura se combina con chips ordinarios de tecnología CMOS (como los de las cámaras fotográficas) para las funciones de ingreso y salida de información, formando la llamada DNN: *Deep Neural Network*. Una CPU de 1 CMOS y 1 chip RPU de 4.096x4.096 puede realizar 51 GOps (giga-operaciones por segundo). Una vez que se pueda construir un computador con una CPU y 100 RPUs podría operar como una red neuronal de 16.000 millones de sinapsis, gastando solo 22 watts. Sería 30.000 veces más eficiente que los sistemas actuales. Ahí apunta actualmente la investigación-desarrollo de IBM. (Extreme Tech, 1/04/2016).

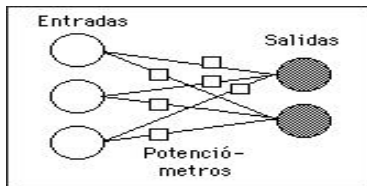
4.2. Redes neuronales

Las redes neuronales artificiales intentan imitar las conexiones de las neuronas cerebrales.

Para comprender las operaciones básicas del tipo más simple de red neural, imaginen tres nodos de entrada dispuestos en una columna y una columna de dos nodos de salida a la derecha de los de entrada. Cada uno de los tres de entrada está conectado por cables a cada uno de los

dos de salida; pero en la mitad de cada cable hay un interruptor que puede cerrar el paso de la corriente eléctrica a través del cable desde el nodo de entrada al de salida.

Cuando se expone cualquiera de los tipos de nodos a una corriente lo suficientemente grande, se dice que el nodo está «activo»; cuando la corriente es muy pequeña o inexistente, está inactivo. Para resolver un problema con una red de este tipo, se puede asignar a cada nodo de entrada la representación de un rasgo particular del problema [y atribuir un determinado significado a los de salida] [...] Por supuesto, la clave para que la red produzca la respuesta correcta está en ajustar los interruptores entre los nodos de entrada y de salida. En las redes más sencillas, este ajuste puede ser simplemente cosa de abrir o cerrar por completo cada uno de ellos (Freedman, pp.90-91).



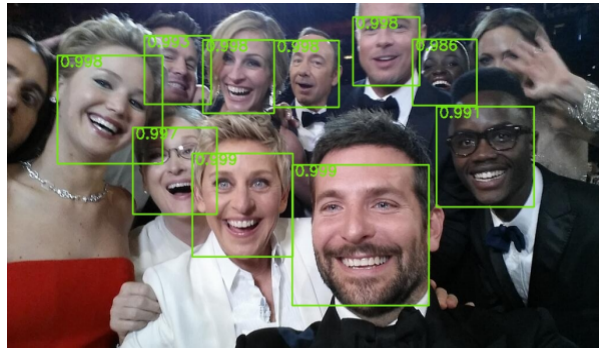
Sin embargo, no se usan hoy simples interruptores (sistema binario) sino sistemas que permiten ajustar los valores en una escala continua (potenciómetros). Se habla así del "peso" de una conexión, el que debe ser ajustado para que una determinada entrada produzca la salida correcta, tarea bastante difícil si no se cuenta con un sistema que permite a la red "aprender" de sucesivas experiencias.

El desafío con estas redes es comprender exactamente qué proceso ocurre en cada capa: "Por ejemplo, la primera capa puede buscar por los bordes o esquinas de los elementos de la fotografía. Las capas intermedias interpretan las características básicas para buscar formas o componentes, como una puerta o una hoja. Y las últimas capas arman ésto en una interpretación completa, las que se activan en respuesta a cosas complejas como edificios o árboles". Pero los resultados pueden ser impredecibles:



Fuente: FayerWayer, 19/06/2015

Lo que Google ha hecho con este tipo de problema es una labor de lo que llaman "inceptionismo": resaltar sobre una imagen convencional las diferencias y detalles que el software reconoce de forma más acusada. De este modo logramos ver algo así como "lo que llama la atención a la inteligencia artificial", algo que normalmente no se ve pero que está dando vueltas por los circuitos de memoria (Microsiervos, 19/06/2015).



Fuente: MIT Technology Review, 16/02/2015

Yahoo y Google utilizan este tipo de redes (simuladas en un supercomputador) para el reconocimiento de rostros. Pero la respuesta no siempre es correcta como lo demostró lo ocurrido en Google al etiquetar como gorilas a dos mujeres de color, por lo que el mismísimo arquitecto jefe de Google+, Yonatan Zunger, debió disculparse, aclarando que el problema no era exclusivo de los afroamericanos, ya que también habían tenido problemas con personas blancas que habían sido etiquetadas como perros y focas, debido a que el aprendizaje de la red de IA era difícil y, por ahora, requiere la intervención de supervisores que "afinen" permanentemente el sistema para descartar los malos resultados (MIT Technology Review, 16/02/2015).

Google también usa su programa *DeepMind* para "enseñar a leer" a sus computadores. Sus redes neuronales "profundas" pueden responder a un 60 por ciento de las consultas formuladas acerca de los artículos conservados (MIT Technology Review, 19/06/2015). Está ahora ejecutando su motor de búsqueda con ayuda del aprendizaje automático, aumentando sus reglas de búsqueda predeterminadas con redes neuronales profundas que pueden aprender a identificar los mejores resultados de búsqueda analizando grandes cantidades de datos de búsqueda ya existentes (Wired, 11/11/2016). Pero

Las redes neuronales profundas sólo funcionan bien en ciertas situaciones, por lo menos en la actualidad. Pueden reconocer las fotos. Pueden identificar las palabras habladas. Ayudan a elegir resultados de búsqueda en Google. Pero no pueden hacer correr todo el motor de búsqueda de Google. Y no pueden ejecutar los *Trending Topics* en Facebook. [...] El punto más grande es que incluso las redes neuronales llevan sesgos humanos. [...] Hay que filtrar la corriente. Y una vez que comience a filtrar la corriente, se hacen juicios humanos. (Wired, 2/09/2016)

Microsoft es otro de los competidores que va en la misma dirección y consigue por ahora una calificación del 29,1%, lo que equivale a que 3 de cada 10 veces sus respuestas son al menos tan buenas como las que podría brindar un humano. La idea es utilizarla para crear subtítulos (pies de fotos). También es el sistema, llamado *Memory Network*, que Facebook está desarrollando para procesar textos y responder a cualquier pregunta acerca de los mismos (MIT Technology Review, 26/03/2015).

Se trabaja ahora en mejorar aún más estos sistemas para llegar a un nivel en que la máquina pueda demostrar su competencia a la hora de analizar una imagen "comprendiendo" qué ocurre en la escena y no solo identificar objetos, o sea pasar de algún modo el test de Turing aplicado a las imágenes. En ello trabaja el profesor Stuart Geman y compañeros suyos de la Universidad Brown, en colaboración con DARPA (Xataka, 11/03/2015).

4.3. Algoritmos evolutivos y aprendizaje automático

Una floreciente vía de investigación en los últimos tiempos es la llamada IA bioinspirada: aprender de los seres vivos. Se crean entornos virtuales en los que se sitúan "organismos" con distintas características para ver cómo evolucionan. La gran idea está en lo que se denomina algoritmo genético o evolutivo. Se propone un objetivo y se ponen a ininidad de «bichos» a intentar conseguirlo. Luego se introduce en cada uno de ellos una variación aleatoria (una mutación). A los que la variación no les mejore para conseguir el objetivo o, incluso, les perjudique, se eliminan. Los que mejoren siguen y se les introduce una nueva mutación aleatoria. Así, tras muchas generaciones, obtendremos seres altamente especializados en lo que queremos obtener. Todo consiste en copiar la selección natural darwiniana, pero acelerarla tanto como nos permiten hoy nuestros computadores (Sánchez-Migallón).

La NASA utilizó ya este tipo de programación para resolver problemas de conexiones diseñando nuevos tipos de antenas para microsatélites en órbita.

Nos encontramos aquí, justamente, en un campo en que parece inevitable introducir nuevas distinciones, como ya lo hacen diversos especialistas. Hace algún tiempo, se empezó a distinguir entre "IA dura" e "IA blanda", la primera apuntando a lograr que una máquina tenga éxito en el test de Turing (tener respuestas indistinguibles de un ser humano en un diálogo a distancia) y la segunda a "aprender" a partir de la acumulación de información y extraer de ello conclusiones útiles. Así, en la "IA blanda" existe un componente de aprendizaje automático ("*machine learning*"). Pero también existen aplicaciones de aprendizaje automático que no implican aspectos de IA. Por ello, Shimon Zilins propone hablar de "inteligencia mecánica" ("*machine intelligence*"). Es la que encontramos en aplicaciones como SIRI, el reconocimiento de caras, las recomendaciones de compra, etc. Es también la que más interesa por ahora, para efectos comerciales, a los "grandes" de internet. Y donde es probable que se multipliquen nuevos productos, más económicos y de más fácil uso para empresas menores (Shimonzilins.com, 10/12/2014).

Sin embargo ya se han dado los primeros pasos hacia máquinas que "piensan por sí mismas". Google anunció un importante paso en esta dirección: su sistema de inteligencia artificial *Google Brain* ha sido capaz de aprender a crear su propio cifrado sin necesidad de intervención humana. Una red neuronal cifra sus mensajes para enviarlos a otra red neuronal y una tercera intenta interceptarlos y decodificarlos. En los primeros intentos, el cifrado era bastante pobre y fueron descifrados, pero después de 15.000 repeticiones la tercera red fue incapaz de descubrir más de la mitad del contenido, una tasa similar a la que se puede obtener por pura casualidad. Lo más extraño (o lo inquietante) de este avance en las redes neuronales es que los investigadores no saben exactamente cómo funciona el método de cifrado exitoso que sus máquinas "imaginaron" (Xataka, 28/10/2016).

4.5. El futuro

Por cierto falta mucho para llegar a robots al estilo de los de Asimov, con una inteligencia casi humana. Por ahora, los esfuerzos de desarrollo de la IA van en otra dirección, conducidos por los intereses económicos. La inteligencia artificial, ya presente en los otros negocios de Google, jugará un papel primordial en sus vehículos autónomos. Para ello, en 2015, contrataron a Geoffrey Hinton, una de las mentes más brillantes en la materia. Apple también estaría trabajando en el desarrollo de su propio modelo de coche y, según Reuters, contratarían hasta 86 personas para incrementar sus efectivos en inteligencia artificial.

Pero ambos se encontrarán con una fuerte resistencia de la industria automotriz, que está llevando a cabo actualmente una verdadera guerra por contratar sus propios expertos en IA. ¿Por qué? Porque también tienen a la vista el desarrollo de los coches autónomos y necesitan buenos sistemas de IA para conducirlos. Y no quieren dejar que sistemas como Android Auto y (Apple)CarPlay acaparen a sus clientes, quedando ellos relegados a simples usuarios de un producto *commodity*. Las soluciones propias les dan mayor independencia. Así, según Bloomberg, Toyota ha contratado a la "cabeza pensante" del equipo de ingeniería de inteligencia artificial del ejército estadounidense, que ayudaría a incorporar en sus vehículos sistemas de seguridad que le permitan reducir el número de accidentes en las carreteras. Está invirtiendo 50 millones de dólares en la investigación de tecnología relacionada con inteligencia artificial y el coche autónomo. General Motors, Ford y Honda también están invirtiendo grandes sumas en I+D. Uber también contrató a 40 investigadores del departamento de inteligencia artificial de Carnegie Mellon, dejando ahí un departamento diezmado. ¿Su objetivo? Con coches autónomos ya no necesitarían a sus conductores, evitando numerosos quebraderos de cabeza (Xataka, 16/09/2015).

Todas las técnicas señaladas son "muy primitivas" a juicio de Rafael Yuste, ideólogo del Proyecto BRAIN¹⁴ (Investigación del Cerebro a través del Avance de Neurotecnologías Innovadoras), cuyo fin último es mapear la actividad del cerebro, lo que permitirá desarrollar técnicas para alterar la actividad de circuitos neuronales y, de esa manera, corregir los defectos que provocan las enfermedades mentales, entre otras utilidades. BRAIN involucra a laboratorios de muchos países, requiere cientos de millones de dólares para ser financiado (una parte de los cuales aseguró el presidente Obama) y la dedicación de investigadores de distintas disciplinas (One, El País.com, 16-09-2015).

Vincent Conitzer, profesor de ciencias de la computación, economía y filosofía en la Universidad de Duke, comparte esta opinión:

Incluso los trabajos más simples requieren habilidades -como la resolución creativa de problemas- para los que los sistemas de AI aún no son competentes.

Es difícil diseñar sistemas AI totalmente autónomos para entornos desordenados y reales. En términos más generales, los sistemas actuales de AI no tienen un amplio entendimiento del mundo, incluyendo nuestras convenciones sociales, y carecen de sentido común.

Los sistemas de AI no son todavía capaces de una verdadera abstracción, de dar un paso atrás, inspeccionar su propio proceso de razonamiento y generalizar lo que está pasando. Una consecuencia es que todavía son limitados en creatividad. Pueden llegar a nuevas soluciones a los problemas: por ejemplo, AlphaGo de DeepMind de Google jugó un movimiento muy inusual en una de sus partidas contra el campeón humano Lee Sedol. Pueden crear algunos tipos de arte, como el arte aparentemente psicodélico producido a través de redes neuronales por DeepDream de Google. Pero no es el tipo de creatividad que realmente da una nueva perspectiva de la situación actual.

Muchos puestos de trabajo seguirán siendo inmunes, al menos en el futuro previsible, ya que fundamentalmente requieren habilidades que son difíciles de replicar en la IA. Considere, por ejemplo, los terapeutas, los entrenadores o los maestros de jardín de infancia: estos trabajos requieren una comprensión general del mundo, incluyendo la psicología humana y el razonamiento social, la capacidad de hacer frente a circunstancias inusuales, etc.

La idea de que el progreso reciente en la IA evitará que la mayoría de la gente contribuya significativamente a la sociedad es una tontería. (MIT Technology Review, 31/10/2016)

Para operar con apariencia de inteligencia, los programas de IA deben contar con el respaldo de enormes bases de datos que contengan toda la información posible acerca del contexto en el cual se hace una pregunta. *"Ninguna tecnología de IA existente puede dominar incluso los desafíos más simples sin un contexto proporcionado por el ser humano. Mientras este sea el caso, la versión de hoy de IA no es realmente «inteligente» y no será la bala de plata para ninguno de nuestros negocios o problemas sociales."* (A.Baciu, Wired, 7/12/2017) En modo alguno representan, en estas condiciones, un riesgo real para la humanidad como han planteado Stephen Hawking y Elon Musk. Y falta muchísimo para que lleguemos a la "singularidad" que pronostica Ray Kurzweil, donde las máquinas inteligentes nos superen (y ni siquiera sabemos si es posible).

5. Después de los Big Data: Smart, Small, Rich

El tema de los *big data* ha invadido ampliamente las publicaciones especializadas relativas a internet y a la informática "en la nube". Pero ya no se trata solamente de hablar de la acumulación de datos recogidos principalmente a partir de las actividades y datos personales de los usuarios de la red, en beneficio de las grandes empresas de internet, sino cada vez más de los servicios derivados y de los beneficios que pueden ofrecer a otras empresas (y a las personas). Incluso ha surgido un enfoque alternativo que nos parece muy prometedor, el de los "*rich data*", aunque es probable que no les guste a los "grandes" (que ya no harían, quizás, tanto "negocio" con nuestros datos).

5.1. ¿Big Data para todos?

Uno de los problemas de los *big data* es la dificultad de su manejo. Los "grandes", como Google y Facebook tienen enormes centros de datos, supercomputadores y departamentos especializados que utilizan inteligencia artificial para extraer información útil para ellos. Muy pocas empresas pueden darse

este lujo. De ahí también que muchos poderosos, como IBM, Microsoft, Google y Amazon, ofrecen este tipo de servicio en su "nube".

Pero aún así no es fácil utilizarlos. Las empresas deben contar con especialistas en la gestión de este tipo de sistema, tanto para ingresar los datos (generalmente en una forma no compatible con los modelos habituales de bases de datos) como para dirigir los procesos a realizar (generalmente mediante programación paralela, algo muy diferente de la programación clásica) e interpretar los resultados. Sin embargo, se ha ido tomando conciencia de esta dificultad y están empezando a aparecer algunas propuestas para facilitar las operaciones.

Así, como ya había mencionado, Microsoft ha lanzado un servicio en la nube llamado *Azure Machine Learning (AzureML)*, un sistema de aprendizaje automático que permite analizar los *big data* para reconocer patrones y extraer valor. Ofrece una interfaz tipo web, de arrastrar y soltar, para colocar fácilmente las piezas del rompecabezas y extraer información sobre la base de conjuntos de datos complejos (Evenbrite.ca, 13/11/2014).

También están surgiendo servicios más puntuales como el de SocialBro, que lanzó "*Monitoring*", una nueva herramienta de minería de datos de Twitter, que tiene como objetivo identificar tendencias de mercado, comportamientos de los consumidores y oportunidades de negocio. La herramienta (de pago), disponible para todos los usuarios, hará posible acceder a datos de Twitter combinando *hashtags*, menciones y palabras clave, obteniendo datos demográficos importantes (género, idioma o localización) así como sentimientos de los tuits (WwwwhatsNew, 10/12/2014).

Apple e IBM, por su parte, han anunciado la primera oleada de *apps* para dispositivos iOS conocidas como soluciones *IBM MobileFirst*. Se trata de una nueva categoría de *apps* y servicios basados en la nube que acercan las herramientas de análisis y *big data* de IBM a los usuarios de iPhone y iPad de las empresas. Ambas compañías afirman que las primeras *apps* diseñadas bajo estas soluciones ya están siendo usadas por Citi, Air Canada, Sprint y Banorte (MacWorld, 10/12/2014). En realidad, no hablan del ingreso de datos, sino del uso, es decir de la consulta de los resultados.

Por ayudar a hacer progresos en este ámbito, el profesor Zoubin Ghahramani y su grupo de la Universidad de Cambridge recibió un premio "*Google Focused Research Award*" en apoyo a su proyecto "*The Automatic Statistician*". Hasta el momento, el proyecto ha sido mayormente centrado en la búsqueda de tendencias en datos de series de tiempo, descubrir automáticamente regularidades e irregularidades, etc. Recurre para ello a modelos estadísticos no paramétricos basados en procesos de Gauss (Google Research Blog, 2/12/2014). Aunque el objetivo declarado de este proyecto es "*construir una inteligencia artificial (IA) para la ciencia de datos*", es un ejemplo de que, para explotar los *big data*, no se requieren necesariamente aplicaciones de inteligencia artificial.

5.2. Smart Data

"*Big data is out, small, smart data is in*" dice David Lavenda, VP Product Strategy de Harmon.ie, un proveedor de productos de apoyo móvil (Cmswire.com, 2/12/2014).

Los "*smart data*" (datos inteligentes) son información entregada para extender nuestros sentidos humanos y conocimiento para que se apliquen a la evaluación y decisión en torno a respuestas a dar, que se combinan con los *big data* para garantizar que podemos actuar sobre la base de un conocimiento completo. Los *smart data* permiten a las personas ser más empoderadas, o inteligentes, y al hacer ésto las hace también conscientes de la necesidad de más datos provenientes de los recursos de *big data* según Andy Mulholland¹⁵. Los *smart data* son datos puestos en su contexto y en las "manos correctas", y relevantes para algunas cuestiones, actividades o problemas específicos. Una vez que los datos se transforman, se procesan y se entregan a la persona adecuada en el negocio que los necesita para hacer algo de valor, pueden ser una parte importante del modelo de detección y respuesta. Pero el proceso de transformación no es ni fácil ni barato (cfr. Michael Fauscette, 2014). No son, por lo tanto, una solución frente a la dificultad de manejar los *big data*, sino un mejor producto del procesamiento de los mismos.

5.3. Small Data

La definición más simple de "*small data*" (datos pequeños) es que se trata de un conjunto de datos que puede caber en un ordenador o dispositivo personal. Pero no basta con considerar el tamaño del conjunto de datos. Hay que considerar también otras dimensiones. Una de ellas es el proceso de tomar grandes conjuntos de datos y distribuirlos a los nodos que harán posible utilizarlos en dispositivos locales

para alguna actividad de negocios. Implica un modelo de computación distribuida del procesamiento de datos y la democratización de la información acumulada, convirtiendo los *big data* en información derivada, procesable localmente, tal vez incluso mezclando varias fuentes y, lo más probable, visualizándolos de alguna manera (interfaz gráfica). Deben ser presentados en forma oportuna, pertinente, organizada y empaquetada para facilitar los procesos de toma de decisión rutinarios (Michael Fauscette, *ibidem*).

Aquí sí surge una pista para abordar los *big data* de otra manera: en lugar de procesar la masa, realicemos un proceso de selección en función de algún criterio inicial (o varios), considerando la utilidad para los futuros usuarios, extraigamos estos datos y procesemos el subconjunto obtenido, lo cual es mucho más fácil (por ejemplo con herramientas de "minería de datos", bien conocidas; y hasta se podrían trabajar, eventualmente, con planillas Excel). El problema se reduce (aunque puede no ser fácil) a elegir los "datos correctos" para las "personas correctas" (*the "right" data for the "right" person*). Dice Fauscette que con esta manera de trabajar es mucho más probable que las empresas contesten mejor a las expectativas de sus clientes.

Ésta es la opción que parece haber tomado Apple para hacer funcionar su "asistente" Siri sin adueñarse de los datos de los usuarios. Craig Federighi, vicepresidente senior de *software* de Apple, asegura que es falso que se deba elegir entre las bondades de la inteligencia artificial y la privacidad de los usuarios. Según él, no hay que elegir entre uno y otro. Lo que aplica la empresa es la llamada "privacidad diferencial", un método estadístico aplicado a la recolección y gestión de datos desarrollado en 2006 por Microsoft Research¹⁶. "Este análisis se realiza de tal manera que no es matemáticamente posible encontrar nada sobre personas individuales. Según Aaron Roth, profesor de ciencias informáticas de la Universidad de Pennsylvania, con este método se consigue impedir que los datos anónimos sean analizados cruzándolos con otras bases de datos." (E. Archanco, Xataka/Applefera, 14/06/2016)

¿Cómo lo aplica Apple? Ha construido un "cerebro" dentro de cada uno de sus nuevos teléfonos, donde se almacenan todos los datos sensibles sobre el usuario, sus hábitos y apps, además de una base de datos sobre información general no relacionada con el usuario. El aprendizaje automático se realiza entonces en cada dispositivo, sin que ningún dato privado se almacene en la nube o pueda ser interceptado durante su transmisión. Por ello, en iOS 10, este cerebro pesa alrededor de 200MB. Es la base de su "asistente" (mayordomo) SIRI.

5.4. Rich Data

El concepto de "*rich data*" fue introducido por Tim Berners-Lee, el prócer de la web. Son datos que "empoderarían" esencialmente al individuo. Según Berners-Lee: "Si un computador recopiló datos de su médico, de su compañía de tarjeta de crédito, de su casa inteligente, de sus redes sociales, y así sucesivamente, podría obtener una visión real de su vida". Éstos serían *rich data* (datos ricos o enriquecidos), que no tienen por que ser acumulados en otra parte sino en su propio PC, ni pasar a pertenecer a grandes empresas. En lugar de recoger todos los datos en un solo lugar (estas empresas) antes de ser utilizados, se pueden mantener en muchos pequeños conjuntos de datos para cada persona y situación. Pero lo más importante es que esta unión de conjuntos de datos debería ocurrir en el punto y momento de una necesidad y de forma dinámica. Este modelo de "segmentación individual" de hecho ya es conocido y existen herramientas para utilizarlo. Su principal ventaja es que solo se necesita un pequeño número de observaciones para responder a las necesidades de un individuo, lo cual es fácil de procesar. Y una persona no necesita en absoluto información acerca de los datos de otras personas: no le son de ninguna utilidad. Es de esperar que aparezcan aplicaciones para que podamos efectivamente hacernos cargos de este modo de nuestros propios datos.

5.5. Propiedad y capitalismo de datos

Esta visión se ha de enlazar con otra: la de la propiedad de los datos individuales. Las empresas que recolectan nuestros datos no son legítimos propietarios de los mismos. Y se ha planteado que deberían pagar por ellos. Por cierto, en muchos casos les pagamos sus servicios aparentemente gratuitos entregándoles nuestros datos. Éste es un modelo de negocio que sería difícil cambiar. Pero es legítimo discutir los límites del mismo y es uno de los temas que abordan las autoridades de varios países como, en particular, la Comisión Europea.

A los sitios de internet y las grandes plataformas (como Facebook, Google, Amazon o Microsoft) se están sumando "corredores de datos" (*brokers*) que crean perfiles personalizados masivos sobre cada uno de nosotros, que luego venden y pueden utilizarse para eludir las reglas de protección de los consumidores

que pretenden limitar las prácticas depredadoras y discriminatorias. Los corredores de datos tienen por objetivo proporcionar sus servicios desde las sombras, amasando miles de millones de datos acerca de las personas en todo el mundo. Y sus bancos de datos nunca les parecen demasiado grandes.

Muchas empresas y gobiernos se sienten impelidos a recolectar la mayor cantidad posible de datos, extraídos de todas las fuentes posibles y de todas las maneras posibles.

Se ha creado una carrera de armamentos para los datos, alimentando el impulso para crear tecnologías de vigilancia que se infiltran en todos los aspectos de la vida y la sociedad. Y la razón de la creación de estas reservas masivas de datos es el valor que puede o podría generar. (Jathan Sadowski, *The Guardian*, 31/08/2016)

De este modo, los datos se transforman en una nueva forma de capital, que las empresas acumulan para mercantilizar y obtener ingresos. Las empresas utilizan los datos para crear valor –para ellas– sin proporcionar nada a cambio a las personas investigadas, que son en realidad los dueños de los datos.

Si bien algunos podrían argumentar que Google y Facebook nos pagan por nuestros datos con servicios «gratuitos», esto todavía no tiene en cuenta la multitud de apropiadores de datos que no tienen intención de proporcionar ningún tipo de beneficio para aquellos cuyos datos obtienen.

Si las corporaciones y los gobiernos van a subir la apuesta por el tratamiento de los datos como un activo, entonces nosotros – los objetivos de este imperio de datos – deberíamos responder del mismo modo. Muchas de las prácticas comunes de recolección de datos en realidad deben ser tratadas como una forma de robo que llamo apropiación de datos – lo que significa la captura de datos de las personas sin consentimiento ni compensación.

Cuando las empresas buscan el consentimiento, es normalmente a través de los términos de los acuerdos de servicio, contratos excesivamente largos llenos de lenguaje jurídico denso que se espera que obtenga el «acuerdo» de los usuarios sin entenderlo. Es una notable victoria para los que se apropian de datos que el allanamiento se ha convertido en el modelo estándar para obtener el «consentimiento». (*ibidem*)

Las tecnologías de vigilancia utilizadas están protegidas por secretos comerciales, lo cual evita que el público sepa qué datos son analizados y cómo influyen en la actividad de sus compradores (e indirectamente de la sociedad). ¡La industria del corredor de datos por sí sola genera alrededor de 200 mil millones de dólares en ingresos anuales! Esto plantea sin duda una cuestión ética pero la situación también debería ser abordada desde una perspectiva jurídica más activa: *"Necesitamos nuevos modelos de propiedad y protección de datos que reflejan el papel que la información tiene en la sociedad"*. (*ibidem*)

Otra forma de capitalismo de datos es la que ejerce Facebook, transformándose en la principal plataforma de distribución de publicidad en internet (Un estudio del Pew Research Center encontró que Facebook es de lejos el sitio más popular para compartir noticias en los Estados Unidos). Haciéndose al mismo tiempo difusor de noticias, ajustando lo que ofrece a cada usuario según los intereses de éste, ninguna empresa periodística es capaz de competirle en materia de financiamiento por publicidad. Ni los sitios web de periódicos más exitosos son capaces de hacer lo mismo, viéndose prácticamente obligados a pasar por Facebook.

La conclusión lógica de este proceso no es sólo la destrucción de los viejos medios de comunicación, sino el fin del periodismo tal como la conocemos.

La «Facebookización» de noticias tiene el potencial de desestabilizar la democracia, en primer lugar, por el control de lo que leemos y, en segundo lugar, mediante la destrucción de los puntos de venta que proporcionan ese material. (Roy Greenslade, *The Guardian*, 20/09/2016)

6. Big data en la prensa¹⁷

La necesidad de acumular y procesar datos es cada vez más ineludible, especialmente para los medios que publican en Internet. Solo las empresas que realmente procesan grandes volúmenes de datos podrán mejorar continuamente sus servicios, descubrir oportunidades y mitigar – en "tiempo real" – los riesgos que significa cualquier cambio en su formato, su selección de contenidos o su forma de

comercialización, como ha sido demostrado por los cambios adoptados por grandes periódicos como The New York Times, Newsweek y Daily Mail, o revistas especializadas como Wired, mientras en España el diario El Mundo lanzó Orbyt, su apuesta en lo que respecta a prensa *online* de pago. No se trata solamente de los llamados *big data*, acumulados por las empresas a partir de las redes sociales para conocer mejor a sus usuarios, programar su publicidad y tomar mejores decisiones tanto en el plano del mercadeo como en la gestión de su producción. Los medios de comunicación se preguntan cómo utilizar mejor la cantidad de datos de todo tipo que entregan otros medios y también los gobiernos y las empresas y que se encuentran en la web. Por ahora, hacerlo no es fácil. Se requiere dominar la llamada nueva "ciencia de datos", y nuevas herramientas de análisis, rol del ingeniero o científico de datos. Cada nueva función descubierta tiene su propia exigencia de competencias y no hay aún ningún "modelo estándar" (si es que llega a haberlo alguna vez).

6.1. Grandes contenidos

En el caso de los contenidos, existen experiencias como la del diario británico Daily Mail, que ha creado "centros de contenidos" asociados a las prioridades de su línea editorial. Su equipo está revisando constantemente los datos disponibles en la red y los analiza para optimizar la experiencia del lector. Busca medir el éxito de los artículos incluso antes de su publicación, para saber qué elementos en la historia misma se pueden mejorar (@cdperiodismo, 28-04-2013). No hay, además, una manera fácil de coordinar los "datos en reposo" (en este caso los contenidos de bases de datos que se publican) con los datos "en movimiento" (en este caso tanto las noticias que particulares publican en las redes sociales como las datos acerca de la conducta de los lectores) señala Kevin Jordan, a pesar de que el éxito de una compañía – particularmente un medio de comunicación – dependa de ello. *"No es poca cosa integrar grandes cantidades de datos de una variedad de fuentes. Es aún más difícil averiguar exactamente cómo traducir esa información en más visitas –en tiempo real–, cliente por cliente"* reconoce Paul Dunay. Es indispensable estar constantemente al día, en "tiempo real", con lo que acontece tanto en las fuentes como en los destinos.

6.2. Datos de los lectores

Medios como The New York Times han aprendido de una primera –mala– experiencia y han asumido esta perspectiva. El Wall Street Journal se está adecuando a las redes sociales con una función específica para que sus lectores creen un perfil personal al estilo de LinkedIn. El diario digital Mediapart, que empezó en 2011, hoy es rentable gracias a una fórmula de destapes y debates, con acceso de pago, que ha dado excelentes resultados. Mirar la gran red y aprender de la conducta y de los intereses de los lectores es clave no solo para sobrevivir, sino para ganar. Para captar y conservar público, la observación de las conductas se está transformando en una pesadilla para las grandes empresas si no cuentan con los recursos necesarios para un procesamiento ágil. Para los que están teniendo éxito, *"el Santo Grial ha sido mirar a la gente en la vida real, mirarla fuera del lugar de trabajo, ver cuando está por su cuenta, ver la comunicación como parte de su curso de vida en general, ¿cómo le ha cambiado la revolución electrónica y en qué?"* dice Kalev Leetaru, un miembro de la Facultad de Biblioteconomía y Documentación de la Universidad de Illinois. *"Eso ha sido algo difícil de ver en el pasado debido a que no ha habido gran cantidad de datos."* Pero ahora hay más datos de lo que la mayoría de las computadoras pueden procesar, procedentes de plataformas públicas de redes sociales como Twitter (Badger, 2013).

6.3. Enfrentar la acumulación

En ambas dimensiones, la empresa requiere contar con aplicaciones especializadas (*software*) y permanentemente actualizadas. La "buena noticia" es que las hay. La "mala" es que el ajuste puede implicar un cambio importante en la forma de operar y en la "cultura interna" de la empresa. Hay que aprender a redefinirse a cada instante, principalmente sobre la base de los cambios de conducta de los destinatarios, que hay que aprender a observar e interpretar. Y, ante todo, hay que tener conciencia de que estos *big data* tienen una importancia fundamental para una empresa de medios (y especialmente para la prensa).

¿Cómo enfrentar estas dificultades? ¿Cómo avanzar? *"Empezar a lo «grande» es empezar la casa por el tejado"* dice Matti Keltanen, experto en diseño de servicios digitales. Hay que saber primero cuales son los datos importantes a acumular, investigar cuales son las herramientas adecuadas para su análisis, juntar una primera cantidad – suficiente para un análisis – y sacar las lecciones de la experiencia. Lo más

probable es que ya se disponga de alguna herramienta adecuada (p.ej. para analizar los datos noticiosos). "El pequeño secreto del Big Data es que ningún algoritmo puede decirte lo que es relevante o desvelarte su significado. Los datos se convierten entonces en otro problema que resolver. El enfoque del Lean Data –usar pocos datos– supone empezar con preguntas relevantes para el negocio y encontrar maneras de contestarlas a través de datos, en lugar de dedicarnos a cribar incontables conjuntos de datos" agrega Keltanen (TicBeat, 19 de mayo 2013).

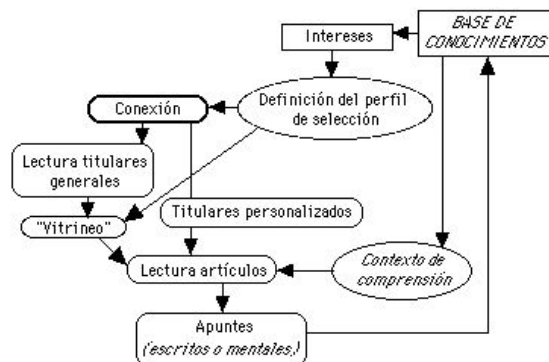
Cuando se crezca en la acumulación de datos, la única vía posible que se visualiza por ahora es la del "aprendizaje de máquina", que hemos descrito antes y para la cual se requieren máquinas poderosas (supercomputadores) o conjuntos de computadores (*clusters*) que se repartan las tareas, algo que se puede obtener contratando servicios en la "nube" de empresas como Amazon, Google, IBM o Microsoft. Sin duda, el software seguirá evolucionando y esperemos que sea en una línea de mayor integración (como ha ocurrido hace años con la ofimática) y con alternativas más compactas y más transparentes, que sean útiles para las PYMES, para los investigadores–académicos y para cada persona deseosa de explorar el nuevo mundo de los datos (y no solo para los más poderosos). Y también con respeto por la privacidad de los datos (que se puede lograr mediante "anonimización", es decir desvinculando y obviando los datos de identidad).

6.4. Servir al lector

Como señalé en una ponencia de 2009, en un periódico destinado principalmente a suscriptores, no es difícil registrar las preferencias de estos por ciertas categorías de noticias, sea pidiendo que las seleccionen al darse de alta sea tomando nota de las que eligen leer a partir de los titulares de la portada común. Obviamente se requiere contar con dos bases de datos conectadas por esta variable común (las categorías de noticias).

La portada con los principales titulares del día y el formulario de identificación permitirá el "vitriero" de titulares y la eventual navegación por las secciones que permanezcan accesibles a todos. El flujo de operaciones para el suscriptor se grafica a continuación (Figura HD1). En este gráfico se observa cómo el "vitriero" puede ser obviado cuando se cuenta con un "archivo de diseminación" que contiene el registro previo del perfil de preferencias de un determinado cliente. Al conectarse e identificarse éste, opera – en forma invisible para él – el proceso de consulta al archivo de diseminación y la consecuente extracción de los titulares que corresponden al perfil personal, para que sean estos los que aparezcan luego en una portada personalizada. Aquí es donde interviene el algoritmo de selección en función del conocimiento del suscriptor que tiene el medio.

Figura HD1: Proceso de lectura de periódico online
(basado en procesos mentales y físicos)



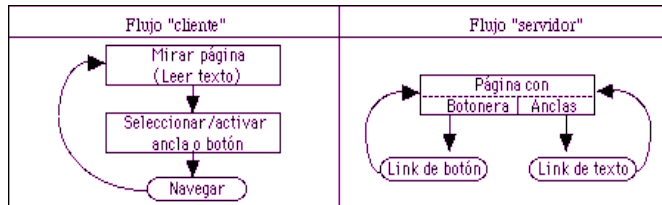
Esto no excluye en ningún caso el menú de acceso a otras secciones-, lo cual dependerá de una definición de una política de difusión propia del medio. En efecto no se trata de negar el acceso del lector al resto de las noticias, ni tampoco – por ejemplo – de mantenerlo en la ignorancia de hechos imprevistos que pueden ser de inmensa importancia para la sociedad en su totalidad y acerca de los cuales podría ser conveniente llamar su atención. ¡Es esencial para un diario digital personalizado asegurar un espacio para sorprender con noticias que se salen de las elecciones personales automatizadas (*big data* o *small data*), como advirtió Andrew Shapiro en *The control revolution* y

Nicholas Negroponte en *Daily Me*. He señalado antes el peligro de reforzar solamente intereses y opiniones personales.

La "Lectura de artículos" y los "Apuntes" señalados en este gráfico no se refieren exclusivamente a los registros digitales directamente asociados a los titulares, sino a toda la navegación que el lector pueda requerir para comprender éstos y ampliar, en consecuencia, su "base de conocimientos". Así, dicha base y el "contexto de comprensión" propios del sujeto son a la vez condicionantes y productos que se enriquecen con la lectura.

Los flujos básicos para el suscriptor y para el servidor se reducen a los simples esquemas de la figura siguiente:

Figura HD2: Flujos básicos de operaciones



A pesar de la simplicidad de estos flujos, el esquema real de navegación de un cliente puede resultar complejo, ya que podría, cuando lo desea, navegar por múltiples archivos y objetos diferentes de una misma gran base de datos. Sin perjuicio de que los archivos señalados en el esquema podrán pertenecer a una misma base de datos relacional orientada a objetos y que las páginas podrán quedar subdivididas (en función de la longitud del original y de los diversos objetos-componentes), el mapa básico de navegación se ajustaría a un esquema de interrelación de archivos propio del tipo de datos y de la estructura de la base de datos que sostendrá el sistema interactivo y que variará según el tipo de empresa u organismo y los fines del proyecto.

7. Poder y ética en la era de los datos

7.1. El cambio del poder

7.1.1. La pérdida del poder tradicional

Como han señalado Alvin Toffler (*"El cambio del poder"*) y mejor aún Moisés Naim (*"El fin del poder"*) un nuevo gran cambio está en marcha ahora: las instituciones, las empresas e incluso las personas (que lo tenían) están perdiendo su poder. Pero no han percibido adonde está yendo el poder, donde está el motor de este cambio: en la explosión de múltiples formas de conexión, no solo en la tecnología, sino también en el comercio, las finanzas, los transportes e incluso la biología, sin que el poder resida ya en una entidad específica. Más que introducir una "era de la información", el cambio actual radicalmente significativo es el desarrollo de una "era de las conexiones".

El poder es ahora de las redes y proviene del número, del tipo y de la velocidad de las relaciones que se establecen. Somos "*hambrientos de conexiones*" dice Cooper (p.33) y la interconexión de redes es tal que llegamos a una "*transición de fase*", un concepto clave para comprender el modo en que un súbito aumento de la conectividad implica una transformación fundamental en cualquier sistema empírico (como ocurre cuando el agua se transforma en hielo y sus moléculas líquidas en cristales).

Los líderes actuales no tienen conciencia alguna de la realidad y del poder de la interconexión mundial en todas las áreas. Viven y tratan de solucionar problemas que son consecuencia de ésta pero no la visualizan en términos obsoletos (Cooper, pp.46-48). Estamos lejos de entender bien la nueva era que empieza: podemos empezar a percibirla pero aún nos falta mucha experiencia.

Las grandes plataformas de hoy (como Google y Facebook) podrían llegar a ser los "porteros" (*gatekeepers*) – y controladores – del sistema mundial, aunque no les interese controlar los contenidos, sino comerciar con ellos. Estos controladores de conexiones son también los que acumulan la mayor cantidad de datos (más que "*big data*" son "*gigantic data*"). Y estamos a punto de multiplicar en forma incalculable su poder con la "internet de las cosas". Quienes controlen ésta "*verán todo, siempre, en todas partes*", obteniendo un poder inmenso (Cooper:87).

7.1.2. Internet como bien público

Como ya señalado, internet fue creada como una red de cooperación entre expertos en informática. No tiene dueño: sus dueños son todos los que ponen sus computadores a disposición de los demás. Requiere sin embargo ser "gobernada", para que funcione adecuadamente, y esto implica definir y actualizar los protocolos de comunicación y administrar el sistema de dominios, lo cual está a cargo de la ICANN (Corporación de Internet para la Asignación de Nombres y Números), organización sin fines de lucro que dejó desde el 1 de octubre de 2016 de ser un organismo bajo el control de los Estados Unidos. Pero gran parte de los servicios que podemos encontrar hoy dependen de las grandes "plataformas": Facebook, Google, Microsoft, Amazon, etc., que no son entidades sin fines de lucro y hacen todo lo posible por controlar la red (como el proyecto de internet gratuita pero minimalista de Facebook en Asia, que solo permitiría el acceso a algunos servicios a través de su plataforma).

"La brecha digital sólo podrá superarse si Internet es vista como un bien público, y los gobiernos deben fijar las normas para involucrar al sector privado", señaló Anita Gurumurthy, de la organización no gubernamental india IT for Change (Informática para el Cambio) (Inter Press Service, 18/11/2005). ¿Pero qué significa "bien público"?

Le pedí a [Lawrence] Lessig que me explicase a qué se refería cuando afirmaba que Internet era un recurso público «común», no dividido entre propietarios privados. Lessig recurrió a la diferencia entre la regulación del ferrocarril y la de las autopistas. En el ferrocarril, los vagones individuales no tienen inteligencia, y sólo un tren puede estar en una vía determinada a una hora concreta, de modo que las redes ferroviarias deben estar muy bien coordinadas desde un órgano central. En cambio, los automóviles tienen supuestamente conductores inteligentes que eligen la ruta adecuada para llegar a su destino sin colisionar con otros vehículos. Ya no se requiere un sistema de coordinación central. «La autopista es un bien común», dijo Lessig. Todo el mundo tiene acceso, nadie necesita una autorización especial para utilizar el sistema de la autopista, cualquiera puede fundar una empresa de transporte por carretera y utilizar la red. Los vehículos que circulan por el bien común de la autopista están regulados: no se puede conducir un tanque, por ejemplo, ni llevar un vehículo sin luces, pues en tal caso el conductor será expulsado. A la luz de la comparación entre el ferrocarril y la autopista, Lessig señaló que «la regulación del espectro podría desplazarse desde el mundo ferroviario, donde los coordinadores centrales tienen que decidir quién utiliza cada vía en cada momento, al de las autopistas, donde los dispositivos inteligentes utilizan el recurso común como quieren». (Rheingold:179)

En una votación esperada desde hace varios años, la Comisión Federal de Comunicaciones de Estados Unidos (FCC) decidió finalmente, tras una larga pugna, clasificar el acceso a la red como un "bien público". Antes lo consideraba como un servicio de información, por lo que carecía de competencia para obligar a las compañías a tratar cada una de las conexiones de los ciudadanos con el mismo principio de igualdad que rige las llamadas de teléfono, para las que no se puede ofrecer un trato prioritario (El País, 27/02/2015). El trato igualitario es lo que se conoce como "neutralidad" de la red e impide a los proveedores de servicios de internet (ISP) de países que adhieren a la neutralidad priorizar cierto tráfico u otorgar gratuidad para algunos servicios (p.ej. acceso a redes sociales sin cargo en la cuenta de datos en los teléfonos móviles), lo cual puede volver más lenta la transmisión para los clientes que no usan los servicios priorizados. Al menos a ese nivel, aunque dependiendo aún de la política de cada país, el acceso a internet debería ser igual para todos. Pero aún así, son múltiples las influencias y presiones que se ejercen, a distintos niveles, sobre proveedores y usuarios (Ver al respecto mi libro "Sistémica de los medios de comunicación").

7.2. El contrapoder de las multitudes

¿Cómo responder a este nuevo desafío? Varios especialistas, tanto teóricos como técnicos están sugiriendo diversas soluciones.

Los lectores de prensa, oyentes de radio y telespectadores eran consumidores, pero los públicos de Internet son «usuarios» con poderes propios. La cuestión más importante en relación con este nuevo enfoque del poder/conocimiento es si crea las condiciones propicias para un contrapoder que sorprendería a Adorno, Horkheimer y Baudrillard, o si se trata de un simulacro más, una simulación de contrapoder que realmente no cambia lo esencial: quién controla los hilos. (Rheingold:223)

7.2.1. Redes distribuidas

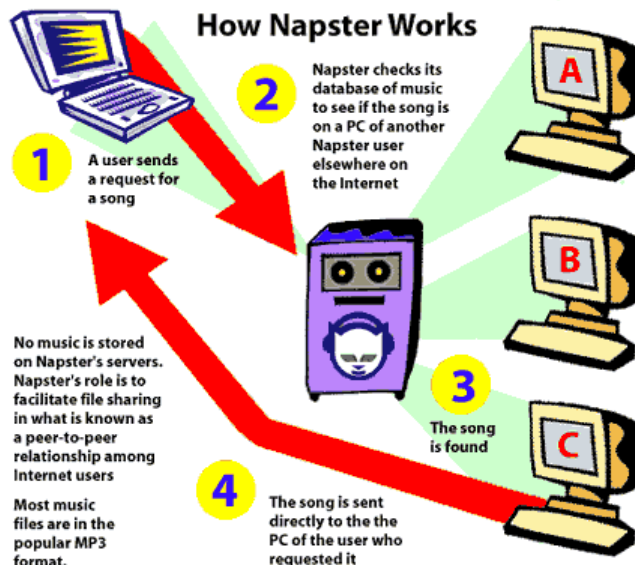
La creación de las primeras redes fue una empresa colectiva, abierta para facilitar el desarrollo colectivo de mejores tecnologías. En un primer momento, se trató de diferentes "terminales" (teclados y pantallas) conectados a un mismo computador central ("servidor"), lo que facilitaba enviar mensajes de un usuario a otro en forma instantánea. Con los PC, se mantuvo el concepto y aparecieron diferentes métodos de comunicación (e-mails, grupos de interés, chats, etc.) pero debían pasar por un servidor que reenviaba los mensajes de un PC a otro.

Paralelamente, sin embargo, se mantuvo la idea de conectar un PC a otro sin pasar por un servidor. Pero, más que para enviarse mensajes, se utilizó para intercambiar archivos: era el P2P (peer to peer) y fue la base del éxito de Napster (1999), que permitía el intercambio de música desde el disco duro de un usuario al disco duro de otro (aunque Napster mantenía una base de datos del contenido disponible). La industria discográfica determinó el cierre de Napster porque el sistema infringía sus derechos. Hubo otros desarrollos que evitaban el uso de un mecanismo central de establecimiento de enlaces y aún hoy opera el P2P, sobre todo para intercambio de archivos pirateados. Pero es un modelo potente, que evita el control por gobiernos y empresas, y es el que proponen Tim Berners-Lee, el inventor de la World Wide Web, y Brewster Kahle, responsable del Internet Archive (Xataka, 8/06/2016).

¿Cómo puede operar sin un servicio central que "sepa" donde se encuentran los contenidos que pueden ser transmitidos? Cuando uno pregunta, por ejemplo, en este tipo de red (como ilustrado arriba),

«Hola, busca recetas de pastel de fresa y ruibarbo», lo que dice en realidad es: «Hola, queridos amigos, ¿podrías decirme si habéis visto alguna receta del pastel de fresa y ruibarbo? Y ya puestos, preguntad a vuestros amigos también y que pregunten ellos también a sus amigos». Es evidente que al cabo de unas cuantas rondas de este tipo, se consigue que un montón de amigos emprenda la búsqueda de la receta. Y es casi imposible que cada cual sepa quién fue el primero en plantear la pregunta.

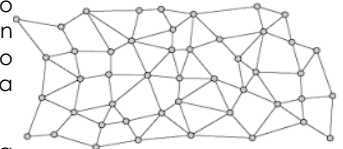
Así que supongamos que un chico del que te separan 6 grados (el amigo de tu amigo de tu amigo de tu amigo) tiene la mejor receta de pastel de fresa y ruibarbo del mundo. Y se lo dice al chico que se lo preguntó, quien, a su vez, se lo dice al que se lo preguntó a él, y así sucesivamente. Al final te llega la respuesta. Pero sólo hay una persona en el mundo que sabe que tú eres quien formuló la pregunta inicial. (GnutelleNet, citado por Rheingold:100)



Fuente: Redeszona.net

Incluso es posible esconder la identidad de esta primera persona.

Se forma entonces una red de nodos que se comunican de uno en uno aunque cada uno puede tener múltiples conexiones. Un nodo, en consecuencia, es una especie de mini-servidor. Esto es un retorno a uno de los principios originarios de la web, como recalca Rheingold y fomenta Tim Berners-Lee.



La compañía MeshNetworks, financiada por la Agencia para la Defensa de Estados Unidos (DARPA), proyectaba ya en 2002 construir un chip de 35 dólares que serviría de punto de acceso inalámbrico. Los teléfonos provistos de este chip podrían funcionar como estaciones de retransmisión (WiFi) para otros teléfonos cercanos, sin pasar por internet. Incluso sería posible con el actual Bluetooth, aunque a menor distancia. (Rheingold:176-177)

Todo, por cierto, depende de la buena voluntad de los participantes y, aquí, entran a jugar las leyes de la cooperación (establecidas en la teoría de juegos). En el caso de la desaparecida red Gnutella, por ejemplo, el 70% de sus usuarios no compartía sus archivos (son los llamados "francotiradores", que quieren recibir pero sin aportar nada), y casi el 50% de los recursos del sistema los aportaban sólo un 1% de los usuarios. Los francotiradores son los peores enemigos de los sistemas cooperativos, lo que ha sido llamado por expertos "la tragedia de los [bienes] comunes". Una solución fue imaginada por la red Mojo Nation [cerrada en 2002]: exigía que cada usuario aportase, al menos, tanto como recibía, asegurando además que nadie supiera dónde estaban almacenados los archivos buscados (*ibidem*:101).

La industria cinematográfica es, como era de esperar, uno de los principales enemigos de este modelo alternativo de internet, abogando por el extremo opuesto: "Hollywood aboga por un nuevo diseño de Internet para detener el intercambio de archivos p2p, lo cual equivale a una propuesta de prohibición sobre el intercambio descentralizado de paquetes en favor de redes centralizadas que se puedan controlar para evitar la infracción de la ley." (Dan Gillmore, Mercury News, 23/03/2002). También pueden oponerse los operadores de cable, al menos para constituir tales redes de PC a PC, ya que pueden filtrar los contenidos y las operaciones realizadas. Pero se encuentran sobrepasados por el predominio, hoy, de las comunicaciones móviles, que pueden ser de aparato a aparato gracias a la tecnología WiFi.

¿Es realmente potente el modelo de red distribuida?

Bob Metcalfe, que dirigió el equipo que inventó Ethernet (conexión por cable) en los tiempos de ARPAnet, calculó que el número potencial de conexiones entre nodos crece más rápidamente que el número de nodos. El valor total de una red en la que cada nodo puede conectarse con todos los demás equivale al cuadrado del número de nodos: con dos nodos, el valor de la unión de ambos es de cuatro unidades; con 4 nodos interconectados, el valor es de 16, y 100 nodos llega a 10.000. (*ibidem*:86)

Pero Metcalfe se quedó corto, como demostró David Reed, de Lotus Corp.. El valor de una red de formación de grupos (RFG) crece aún más rápidamente: la fórmula es dos elevado a la potencia del número de nodos, en lugar de elevar este número al cuadrado. Así, mientras el valor de 10 nodos es 100 (10 elevado al cuadrado) en la ley de Metcalfe, es de 1.024 (2 elevado a 10) en la ley de Reed.

Esto explica que las redes sociales, creadas por correo electrónico u otras comunicaciones sociales, hayan expandido el crecimiento de la red más allá de las comunidades de ingenieros, hasta el punto de incluir toda clase de grupos de interés. La ley de Reed es el enlace entre las redes informáticas y las redes sociales. [...]

Cuando una red pretende aportar algo de valor a los individuos, como una red de televisión, el valor de los servicios es lineal. Cuando la red permite transacciones entre los nodos individuales, el valor se eleva al cuadrado. Cuando la misma red incluye procedimientos para que los individuos constituyan grupos, el valor es exponencial. (Rheingold:87)

Y el aumento de escala fomenta el desarrollo de nuevas aplicaciones que pueden ser "rompedoras" ("killer apps") y encontrar eventualmente un éxito fulgurante, capaz de transformar las relaciones en red.

Así, no es iluso pensar en una nueva internet basada en la fórmula P2P y liberada de la tutela de plataformas comerciales. Alguna coordinación puede ser necesaria, pero lo ha logrado ICANN con el sistema de nombres de dominio, y BitTorrent, quizás la aplicación más extendida para intercambios de archivos y que incluso compañías como Warner Bros o la BBC, empezaron a utilizar como una alternativa a la distribución convencional de películas y programas de televisión.

Para que este tipo de red funcione bien para la transmisión de archivos, se necesitan computadores con buenos discos duros (hoy 500GB o más) y que estén encendidos ojalá todo el día. No todos estarán dispuestos a mantenerlos encendidos y para que los móviles se conviertan en nodos útiles les falta tanto más memoria como planes de datos más económicos (y baterías de mayor duración). La gran pregunta es si las telecoms los permitirán y lo facilitarán o tratarán de enriquecerse cobrando por este tipo de intercambio.

Mientras estén encendidos y conectados, los PC tienen un gran potencial (ciclos de procesador) desaprovechado ("tiempo libre") y si nuestro disco duro contiene cosas que nos interesan ¿por que no facilitarlas a otros interesados? No se trata de dar acceso a todo: podemos reservar una carpeta para los intercambios. Y no faltan los mecanismos para evitar el acceso de indeseables.

Numerosas compañías están visualizando desde hace varios años las potencialidades de este tipo de red y hacen experimentos con redes distribuidas internas (como es el caso de Pfizer, Ericsson, Hitachi y BMW, por ejemplo). Apple, sin embargo, parece preferir otro camino (que todos pasen por sus propios servicios), como indicó su decisión de dejar de fabricar sus enrutadores WiFi Airport (ComputerWorld, 21/11/2016).

Con las tecnologías inalámbricas se abren nuevos modos de concebir la conexión de los usuarios de Internet a los servicios de información, pero lo que se debe preservar ante todo es el derecho a innovar en ese terreno de los procedimientos de conexión a la red. El derecho de conectar todo tipo de dispositivos a la red para hacer cosas que nunca imaginaron sus arquitectos garantizará un amplio margen de innovación en torno a la Internet móvil. ¿Avanzamos hacia un mundo inalámbrico controlado, donde el equivalente a las compañías telefónicas o por cable decida lo que podemos hacer con nuestros aparatos móviles? [...] ¿O adoptaremos una arquitectura para sistemas inalámbricos donde nadie decida las tecnologías que se pueden utilizar y las que no? Cuando construyamos un bien común de innovación, creo que veremos la siguiente gran revolución en la tecnología de Internet inalámbrica. (Lawrence Lessig, citado por Rheingold:83)

7.2.2. Cooperación

A pesar de estas propuestas, el problema con el que nos encontramos enfrentados por ahora está relacionado con las pretensiones de las grandes plataformas y de los operadores de telecomunicaciones de sustituir el libre acceso y libre uso por "sistemas vallados", como denuncian tanto H. Rheingold como J. Cooper. No solo se trata de exigir un registro previo, con entrega de datos personales, sino de alterar incluso la arquitectura del sistema. Además, las tecnologías inalámbricas abren nuevos modos de concebir la conexión de los usuarios y la posibilidad de crear y operar redes exclusivas, limitando no solo el uso por parte de los clientes sino también el derecho a innovar, a hacer cosas que nunca imaginaron los arquitectos de la red.

Sin embargo, aún en este ambiente, es posible utilizar los servicios de la red para contrarrestar esta tendencia al control. No debemos olvidar que, como lo ha hecho notar la teoría de las redes sociales, los nodos de estas redes son siempre seres humanos individuales, aún cuando se consideren las relaciones entre organizaciones: "lo colectivo es *construido* o se reduce a un conjunto de relaciones entre estos individuos" (Callon y Ferrary:9). Un claro ejemplo es el de las redes de intercambio de conocimientos, que son más que enlaces en internet e incluso que relaciones entre personas:

Las verdaderas redes generadoras de conocimiento están en otra parte, reúnen a académicos, universidades, conferencias, servidores, libros, normas y prácticas, los fenómenos que estudian y las herramientas y laboratorios que les permiten hacerlo. «Internet» puede fortalecer y, en ocasiones, debilitar algunas de estas redes —y sin dudas crea las condiciones para que surjan otras nuevas— pero no genera un cambio fundamental en lo que cuenta como conocimiento ni en el modo en que se produce. (Morozov:58)

Se trata claramente de un tipo de red cooperativa, que podemos encontrar en muchos otros ámbitos, donde internet se transforma en una herramienta que ayuda a la cooperación. Y aquí entran a jugar las reglas generales de la cooperación:

Las comunidades modestas salen más beneficiadas si eligen cooperar que si se someten a los poderosos. Y aún más: los débiles de una red de redes tienen en sus manos el destino del conjunto. Lo prueba un modelo matemático desarrollado por tres investigadores españoles. Las

redes de microcrédito en la India, la colaboración entre grupos de investigación y la conservación del oso pardo en los Pirineos son solo algunos de los ejemplos. «Las matemáticas nos dicen que la cooperación suele ser la mejor estrategia para los débiles y que, curiosamente, es también la más beneficiosa para el conjunto. Además, cualquiera de los débiles en solitario puede forzar esta situación. Un poder del que carecen los fuertes» dicen Jacobo Aguirre, investigador del Centro Nacional de Biotecnología del CSIC (CNB-CSIC), Jaime Iranzo, de los Institutos Nacionales de Salud de EEUU y Javier M. Buldú de la Universidad Rey Juan Carlos y el Centro de Tecnología Biomédica (UPM). Estos autores combinaron en su estudio la teoría de juegos, conocida por los trabajos del Nobel John F. Nash y la ciencia de redes complejas. «Cualquiera de los competidores débiles tiene el poder de inducir un cambio de estrategia global sin importar lo que haga el competidor fuerte, que es incapaz de cambiar la situación. Este es el gran poder de los indefensos. Tienen en sus manos el destino del conjunto», explica Javier M. Buldú¹⁸. (Agencia SINC, 14/10/2016)

¿Pero cual es la base de la cooperación?

Me dirigí a los sociólogos, evolucionistas y teóricos del juego, y hete aquí que emergió la reputación como el único hilo conductor que enlazaba la misteriosa generosidad de los cazadores en Tanzania, el placer peculiar que produce el castigo de los timadores, la función social del rumor, la posibilidad de que el lenguaje evolucionase a partir de conductas de limpieza y el modo en que controlan algunas comunidades sus bienes comunes sin provocar tragedias. En cada uno de estos ejemplos, la reputación es el ingrediente secreto de la cooperación.

Las estrategias cooperativas como «Tit for Tat»¹⁹ triunfan porque indican una voluntad de cooperación, si bien se defienden de la explotación a través de represalias contra la no cooperación. Estas dos sencillas estrategias, tomadas en conjunto, explican que los individuos con intereses privados acuerden cooperar en beneficio de la colectividad en muy diversas situaciones. Los organismos en los que se han observado conductas de cooperación, desde los espinosos hasta los vampiros, parecen actuar según un principio de reciprocidad: ofrecen cooperación mutuamente beneficiosa sólo a los compañeros dispuestos a devolver el favor; en cambio, castigan a aquellos que en el pasado no actuaron con reciprocidad, negándose a cooperar con ellos ahora. Es lo que se conoce como «altruismo recíproco». (Rheingold:154)

Las redes distribuidas funcionan bien, evidentemente, en este tipo de ambiente de cooperación. Los "nodos" no cooperantes quedarán excluidos y se verán obligados a buscar otra forma de conexión.

Cuanta más gente descubra que puede obtener beneficios personales si invierte en confianza y practica la cooperación, más invertirá en la empresa cooperativa y ayudará a los demás a sumarse a la aventura. (Rheingold:238)

Robert Wright también observó que

La selección natural, a través de la evolución del «altruismo recíproco», ha incorporado en nosotros varios impulsos que, por cariñosos y sensibleros que parezcan, están diseñados para el fin frío y práctico de aportar un intercambio beneficioso.

Entre estos impulsos se encuentran los siguientes: la generosidad (si bien selectiva y a veces cautelosa); la gratitud y un sentimiento de compromiso diferido; así como una creciente empatía y confianza hacia quienes demuestran que actúan en reciprocidad (también llamados «amigos»). Estos sentimientos, y las conductas que fomentan de forma fructífera, se dan en todas las culturas. Y el motivo, según parece, es que la selección natural «reconoció» la lógica de la suma no cero antes de que la reconocieran los humanos. (de su libro *Nonzero*:22-23, citado por Rheingold:239)

7.2.3. La inteligencia colectiva

El 20 de enero de 2001, el presidente de Filipinas, Joseph Estrada, se convirtió en el primer jefe de Estado de la historia que perdió el poder como producto de la movilización de más un millón de residentes de

Manila, coordinados mediante mensajes de texto (Rheingold:183). Este tipo de movilización es conocida como "estrategia de enjambre" y ha ocurrido también en manifestaciones políticas en los Estados Unidos (Seattle y San Francisco), Senegal y Gran Bretaña, así como las más conocidas de países del norte de África ("primavera árabe").

Los teléfonos móviles adquirieron entonces no sólo el poder de superar las condiciones de masificación y los entornos congestionados como consecuencia de la incapacidad del Estado para ordenar la vida cotidiana, sino que además dieron origen a un nuevo tipo de multitud, una multitud con conciencia propia de movimiento dirigido hacia un objetivo común." (Vicente Rafael²⁰, citado por Rheingold:183)

Pero, como señaló Mark Granovetter, los vínculos débiles (como los que hay entre quienes siguen a algunos líderes en las redes sociales) pueden generar tanto una "coordinación social" como un movimiento de opinión que se queda sin lograr su fin (Sey & Castells, 2004).

Para Rheingold, la manifestación de Manila habría sido una de las primeras manifestaciones de "multitud inteligente". ¿Pero se puede realmente hablar de colectividad inteligente en estas movilizaciones, ahora conocidas como "smart mobs" y "flash mobs"?

La inteligencia de las masas, inteligencia colectiva o sabiduría de los grupos es un concepto que ha sido presentado y desarrollado en 2004 por James Surowiecki²¹. De los casos estudiados, concluyó que la combinación de la información en grupos termina en decisiones que son a menudo mejores que las que podrían haber sido tomadas por un solo miembro del grupo. Uno de los ejemplos que pone en su libro es cómo un gran grupo de gente consigue averiguar con bastante precisión el peso de un buey. Al calcular el promedio de las estimaciones hechas por cada uno, se descubrió que éste estaba más cerca que la mayoría de las estimaciones por separado, incluso que cualquier estimación de los expertos en ganado. Pero este fenómeno ocurre solamente si se cumplen varias condiciones, que serían las siguientes:

Diversidad de opinión: cada persona debería tener información privada aún si es sólo una interpretación excéntrica de los hechos conocidos.

Independencia: las opiniones de la gente no deberían ser determinadas por las opiniones de los que los rodean.

Descentralización: la gente debería poder especializarse y recurrir al conocimiento local.

Combinación: existen algunos mecanismos para convertir los juicios privados en decisiones colectivas. (Wikipedia)

Otros estudios muestran que es importante que haya un lugar donde las opiniones puedan agregarse (lo cual parece obvio) y que haya incentivos (es decir que se pueda ganar algo al acertar). Por el contrario, la inteligencia colectiva falla cuando hay demasiada centralización, demasiada división y demasiada imitación.

La teoría de Surowiecki se está aplicando principalmente en el mercado de predicción (un mercado especulativo o de apuestas creado para realizar especulaciones verificables), pero parece conveniente tener a la vista este concepto y sus mecanismos cuando se plantean los posibles aportes de las redes colaborativas cuya constitución se ve facilitada por los nuevos medios.

Rheingold, siguiendo al ya citado Robert Wright, asegura que "*El incremento del número de personas que participan en juegos más complejos de suma no cero²² produce efectos emergentes como ciudades vibrantes, cuerpos de conocimiento, obras maestras arquitectónicas, mercados y sistemas sanitarios públicos.*" (238)

Bernardo Huberman consideraba útil concebir la inteligencia que emerge en los grupos como una "computación social":

"La inteligencia no se limita al cerebro; también surge en los grupos, como en las colonias de insectos, en la conducta social y económica de las sociedades humanas, así como en las comunidades científicas y profesionales. En todos estos casos, los numerosos agentes capaces de desarrollar tareas locales, que pueden concebirse como computaciones, desarrollan una conducta colectiva que consigue resolver muchos problemas que trascienden la capacidad de cualquier individuo. [...] Cuando interactúan numerosos agentes capaces de realizar procesamiento simbólico, aparecen nuevas regularidades universales en su conducta global. Además, tales regularidades son cuantificables y pueden probarse experimentalmente." ("*The social mind*"²³, citado por Rheingold:205)

Podemos también acercarnos a la problemática de la inteligencia colectiva desde otra perspectiva, como hizo Steve Johnson: estudiando la emergencia de la inteligencia a través de la evolución, en los animales sociales²⁴. Pero adentrarnos en este tema nos alejaría del propósito de este libro.

7.2.4 Los riesgos de los algoritmos “inteligentes”

Pero esta “inteligencia grupal” puede también llevar a grandes errores, del mismo modo que la “inteligencia” de los asistentes personales de Apple, Amazon, Google y Facebook. Ya he señalado como los algoritmos distorsionan el conocimiento al concentrar la información seleccionada en torno a nuestros gustos y creencias, pero también de acuerdo a los intereses de la plataforma de los provee. Y podría llegar de este modo a distorsionar el debate público.

“Nuestro mayordomo [asistente digital] puede recomendar servicios y productos que fomenten los intereses financieros de la super-plataforma en lugar de nuestros propios intereses. Al servir a sus verdaderos amos -las plataformas- puede distorsionar nuestra visión del mercado y llevarnos a los servicios y productos que sus amos desean promover.

El aumento en el poder económico de la super-plataforma puede traducirse en poder político. A medida que dependemos cada vez más de uno o dos mayordomos de cabecera, la super-plataforma aprenderá sobre nuestras creencias políticas y tendrá el poder de afectar nuestras opiniones y el debate público.” (M. E. Stucke y A. Ezrachi, *Wired*, 29/11/2016)

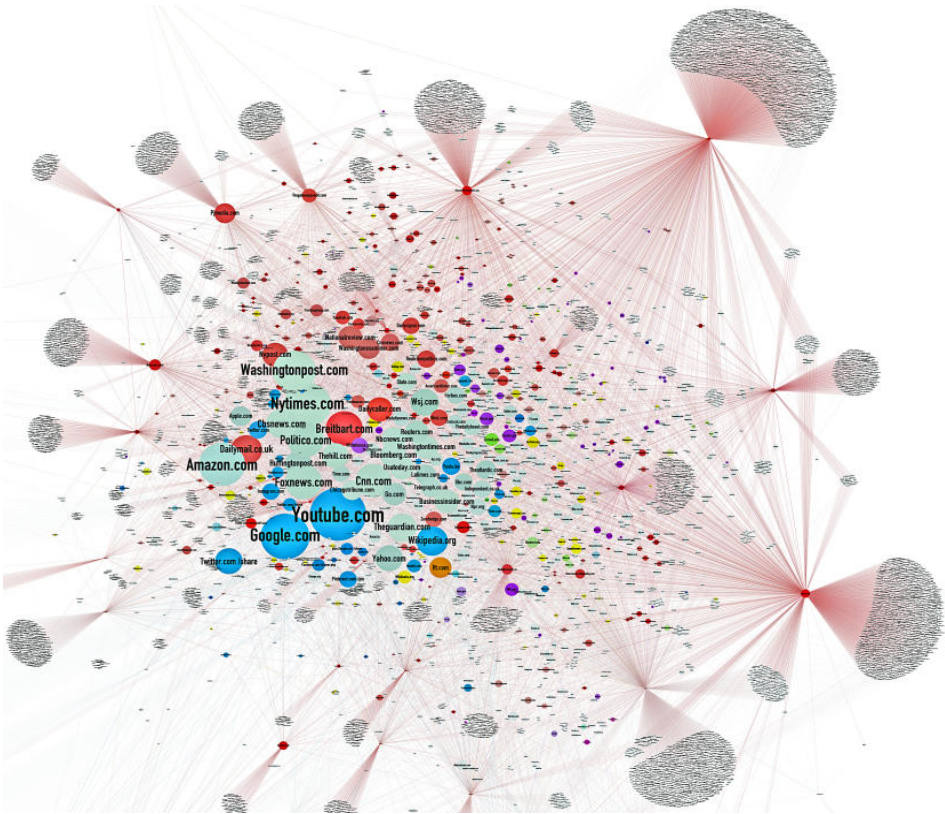
En 2012, Facebook realizó un experimento manipulando el “news feed” de una importante cantidad de sus usuarios, sin su conocimiento, para descubrir como transmitían emociones positivas y negativas. Cuando redujo el contenido positivo, las actualizaciones de los lectores se volvieron menos positivas. Lo mismo ocurrió con contenido negativo. “Al felicitarnos y alentarnos, alentándonos a comunicarnos con los demás, y enviando notas personalizadas en nuestro nombre, puede afectar potencialmente nuestros estados de ánimo y los de nuestros amigos” agregan Stucke y Ezrachi. Es una de las razones por las cuales existe un botón “Me gusta” pero no un “No me gusta”: Facebook no quiere alentar la negatividad.

Filippo Menczer, director del Center for Complex Networks and Systems Research de la Universidad de Indiana (Bloomington) ha estudiado este fenómeno, ligado también a la industria de las noticias falsas:

Tenemos una industria de falsas noticias y desinformación digital. Los sitios de Clickbait fabrican bromas para ganar dinero con anuncios, mientras que los llamados sitios hiperpartidistas publican y difunden rumores y teorías de conspiración para influir en la opinión pública.

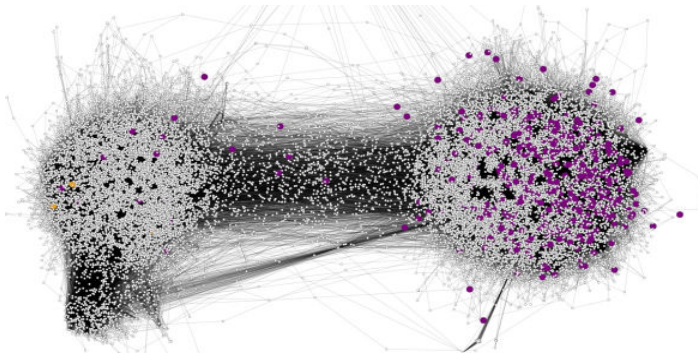
Dentro de nuestra burbuja, estamos expuestos selectivamente a información alineada con nuestras creencias. Ese es un escenario ideal para maximizar el compromiso, pero uno perjudicial para desarrollar un escepticismo saludable. El sesgo de confirmación nos lleva a compartir un titular sin ni siquiera leer el artículo (Menczer, *Inverse.com*, 28/11/2016).

Su investigación ha mostrado que es inevitable que algunos memes se vuelvan virales, independientemente de su calidad, aún cuando los individuos seleccionen información de mayor calidad. Basta prestar más atención a algunos tópicos para que se difunda más información al respecto. Pasamos a pertenecer a una “cámara de eco”, donde rebotan y se multiplican los mensajes acordes con nuestros gustos y creencias, lo cual demostró empíricamente (ver gráfico).



Al centro están los principales medios de prensa y, alrededor, los rebotes y las falsas noticias relacionadas. (De Menczer).

El físico teórico Cristián Huepe, investigador del Instituto de Sistemas Complejos de la Northwestern University en Chicago, también estudió la dinámica de las redes de opinión, comprobando matemáticamente que el algoritmo de sitios como Facebook, sumado a la tendencia de las personas a asociarse con quienes piensan del mismo modo, genera dos grupos que no conversan entre sí.



Fuente: Menczer.

Ese algoritmo diseñado inicialmente para que uno pase más tiempo en Facebook, tiene como consecuencia secundaria que refuerza una situación ya natural en los seres humanos: que uno solo ve lo que a uno le gusta, con lo que está de acuerdo. Y refuerza esa tendencia tribal a separarse en grupos con opiniones muy distintas.

Llega el punto en que las realidades, entre comillas, que tiene cada grupo son distintas. Porque la información que reciben solo está de acuerdo con ellos, entonces interpretan las noticias o incluso repiten informaciones falsas de manera que nadie cuestiona, porque todos ya están de acuerdo con la conclusión (Huepe, El Mercurio, 26/11/2016).

Según un análisis de Buzzfeed, durante la campaña presidencial norteamericana, las 20 noticias inventadas más compartidas en Facebook superaron ampliamente a las 20 noticias verídicas más compartidas (8,7 millones de interacciones, frente a 7 millones). *"Nadie nunca chequea los datos. Así fue como Trump fue elegido. Él dijo lo que quiso, y la gente creyó todo, y cuando se comprobó que las cosas que decía no eran verdad, a la gente no le importaba, porque ellos ya lo habían aceptado. Es tenebroso"*, confesó Paul Horner, uno de los autores más prolíficos de noticias falsas que se viralizan por las redes sociales y con las cuales gana hasta US\$ 10 mil al mes por la publicidad que obtienen sus artículos mentirosos (El Mercurio, 26/11/2016). Vuelvo sobre este tema en el apartado 7.4.4. ("El mal uso de los algoritmos").

7.3. Inteligencia humanística

Ya he mencionado las propuestas (especialmente de Tim Berners-Lee) de una internet descentralizada, que pueda escapar del control de las grandes plataformas y de los gobiernos interesados en controlarla en algunos países.

Quiero ahora dar a conocer un manifiesto que data de 2001 y va en el mismo sentido, pero con un fondo "filosófico" muy interesante. Se trata de una propuesta de Steve Mann²⁵, conocido como padre del "computador ponible" (*wearable*), en ese tiempo en el MIT y hoy profesor de la Universidad de Toronto. Si todos llevamos sensores (y cámaras), somos dueños de lo que captamos y estos datos no deberían ser captados por terceros sin nuestra autorización (la que solo daríamos con muy buenas razones).

El humano dota a la máquina de mucha más inteligencia de la que tendrá la llamada Inteligencia Artificial, y la máquina proporciona al humano capacidades nuevas de manipulación intelectual de la información visual. En cierto sentido el humano usa al ordenador como periférico, y el computador usa al humano como periférico, y los dos forman un solo dispositivo, un bucle de realimentación en el que cada uno extrae del otro su inteligencia. Esto es lo que yo llamo Inteligencia Humanística (Entrevista en El Mundo, 27/12/2000).

Como nosotros somos quienes manejamos los datos, somos nosotros quienes les damos sentido, que hacemos de estos aparatos máquinas "inteligentes". Para Mann, la tecnología actual de redes tiende a "encapsularnos", principalmente en el mundo de la publicidad y el mercadeo, y la mejor forma de reaccionar es utilizando una tecnología de mediatización personal de la realidad que filtre la entrada y salida de información. Con los aparatos portados, con los sensores capaces de comunicarse, podemos hacer que la gente sea más inteligente, dice. Y es más importante que mejorar las herramientas.

En un marco de IH [Inteligencia Humanística], el objetivo es mejorar la inteligencia de la especie, no sus herramientas. Gente inteligente significa, sencillamente, que debemos confiar en la inteligencia humana para el desarrollo de infraestructura tecnológica, en lugar de excluir al ser humano de la ecuación.

Un objetivo importante de la IH es dar un primer paso hacia un principio fundamental de la Ilustración, la dignidad del individuo. Esto se logra, metafórica y literalmente, a través de una transformación protésica del cuerpo en un espacio soberano, permitiendo que cada individuo controle el entorno que le rodea. [...] Uno de los principios básicos del desarrollo tecnológico en el sistema IH es que el usuario debe ser parte integrante del circuito discursivo. (Mann y Niedzviecki, en Rheingold:133)

Mann aboga por la conformación de comunidades de "ciberorganismos" (humanos con recursos cibernéticos, no robots) que denomina «espacio ciborg». No se trata, a su juicio, de una distopía deshumanizada, sino de una estrategia defensiva contra la tiranía tecnológica.

7.4. Rol de la ética en sistemas automáticos

7.4.1. Acuerdos científicos

En la tarde del primer domingo de 2015, Elon Musk subió al escenario en una conferencia a puerta cerrada en un centro turístico de Puerto Rico para discutir acerca de la *"explosión de la inteligencia artificial"*, que se está dando en el campo de la capacidad cognitiva de los programas de IA. Se le unieron Stephen Hawking y expertos de varios laboratorios de IA para discutir sobre los riesgos que esto podría acarrear para la raza humana. A esta conferencia, titulada *"El futuro de la IA: Oportunidades y Desafíos"* (*"The Future of AI: Opportunities and Challenges"*²⁶), concurren importantes académicos como el especialista en ética Nick Bostrom, de Oxford, junto con ejecutivos de la industria como el fundador de Skype, Jaan Tallinn, y expertos como Shane Legg, de Google AI. *"Hay una aceleración en el desarrollo de sistemas de inteligencia artificial"*, dijo en el evento Bart Selman, profesor de Cornell y experto en ética de IA. *"Y ésto es lo que hace más urgente analizar este asunto."*

Al final de la reunión, los delegados firmaron una carta abierta comprometiéndose a llevar a cabo la investigación en IA para el bien y *"evitar los peligros potenciales"*. Todos estuvieron de acuerdo en que la seguridad de la IA es importante. *"Recomendamos más investigación enfocada a asegurarse de que los cada vez más capaces sistemas de inteligencia artificial son robustos y beneficiosos: nuestros sistemas de inteligencia artificial deben hacer lo que nosotros queremos que hagan"*, afirmaron.

Entre los presentes estaban nueve investigadores de DeepMind, la compañía de IA que Google había adquirido. Ellos también han firmado la carta de compromiso.

Entre las preocupaciones expresadas en Puerto Rico está evitar que los robots arruinen la economía o, en general, se multipliquen sin freno. O lleguen a perjudicar a los humanos. Cuatro factores deben ser especialmente considerados:

Verificación: demostrar que el sistema se ha construido de acuerdo con las instrucciones deseadas

Validez: asegurarse de que un sistema que respeta las propiedades formales de diseño no tenga comportamientos no deseados

Seguridad: prevenir la manipulación deliberada por parte de terceros no autorizados

Control: permitir el control humano significativo sobre un sistema IA después de iniciadas sus operaciones.

Un problema típico sería que *"si una IA selecciona las acciones que mejor le permiten completar una tarea, entonces el evitar las condiciones que impidan que el sistema siga ejecutando dicha tarea es un objetivo natural para ella"*, lo cual podría llevarla a actuar contra los humanos, de no haber un mecanismo de control que permita la interrupción. Es lo que había previsto Isaac Asimov hace más de 70 años, cuando propuso sus *"Tres leyes fundamentales de la robótica"* (por primera vez en el relato *"Runaround"*, de 1942, *"El círculo vicioso"* en español). La Ley n°1 es *"Un robot no hará daño a un ser humano o, por inacción, permitirá que un ser humano sufra daño"*.

Los acuerdos de Puerto Rico parecen ser los primeros pasos hacia la instauración de las leyes de Asimov. Pero algunas empresas ya se habían adelantado. El año pasado, los expertos en robótica canadienses de Clearpath Robotics prometieron no construir robots autónomos para uso militar. Google debe asegurarse de que sus autos sin conductor no matarán a los peatones. Y también plantearía un problema ético el que una tienda *online* modifique sus precios para diferentes "nichos" de compradores en función de una u otra característica de su gusto.

Pocos días después del encuentro, Elon Musk regaló diez millones de dólares para iniciar los estudios requeridos.

7.4.2. El problema de los objetos conectados

Con los objetos conectados y las máquinas "inteligentes", las personas delegan o delegarán cada vez más decisiones en la programación de estos aparatos. ¡Llegó la hora de pensar en la importancia de la ética en la robótica! Y recordar que Isaac Asimov ya había pensado en ello y formulado en 1942 sus "leyes básicas de la robótica", que debían quedar impresas en el cerebro "positrónico" de todos sus robots.

Elon Musk cree que los coches conducidos por humanos acabarán siendo ilegalizados en favor de los que están conducidos de forma autónoma. ¿La razón? Las máquinas pueden conducir de forma mucho más segura y los seres humanos acabarían siendo un mayor riesgo en las carreteras. El cambio, por cierto, tomará tiempo: Musk calcula 20 años (Xataka.com, 18/03/2015).

Las primeras pruebas han dado buenos resultados y quedan poco problemas técnicos por resolver, no siendo el menor el de la seguridad informática (Es actualmente imposible descartar toda posibilidad de hackear el sistema y tomar a distancia el control del auto). Pero surge un problema mayor: el ético, implicado en la programación de las decisiones en situaciones complejas. El problema no es tan simple como aplicar las leyes de Asimov. Las situaciones reales son generalmente complejas y suponen múltiples ponderaciones. ¿Que hacer si un niño cruza bruscamente la carretera y evitarlo significa chocar contra un árbol u otro vehículo, poniendo en riesgo la vida del o los pasajeros? ¿Contar la cantidad de posibles víctimas? ¿Tener en cuenta la edad de cada uno?

No parece haber duda que un sistema automático sería más seguro que un conductor humano ya que no se distraería y podría controlar todas las variables deseables... ¿Pero podría considerar mejor los posibles imprevistos y ponderar adecuadamente una respuesta?

¿Pero que pasará con todos los datos personales que podrían estarse "cruzando" con los de los *gadgets* de la internet de las cosas? ¿Dejaremos, por ejemplo, que nuestro refrigerador ordene al supermercado los alimentos que nos faltan considerando nuestro estado de salud (informado por nuestro reloj o pulsera de *fitness*) y las recomendaciones que nuestro médico introdujo en nuestro historial?

Al menos algunos laboratorios se preocupan, como Kaspersky Lab, que se ha unido a la comunidad de bio-hacking sueca BioNyfiken para analizar los problemas de seguridad involucrados²⁷.

7.4.3. Los servicios e implantes personales

La inteligencia artificial usada crecientemente en el diagnóstico de situaciones complejas de salud sufrirá de problemas similares: ¿que malestar resolver primero? ¿Hasta donde son aceptables ciertos efectos secundarios? Aunque es probable que aquí las juntas médicas se sigan imponiendo por mucho tiempo.

Pero también hay que tomar en cuenta que la IA se irá introduciendo en los cuerpos a través de los implantes de todo tipo (ya los hay para controlar el Alzheimer y el Parkinson, y los chips de identificación). Se proponen algunos²⁸ para controlar la ansiedad que podrían ajustar la música ambiental o seleccionar un programa de televisión que nos relaje. Pero, si están abiertos a la red, los publicistas también se verán atraídos, para orientar nuestras emociones. ¿A que otros riesgos nos expondrían? ¿Un hacker podría también cambiar nuestras emociones o el contenido de nuestra "memoria adjunta" (artificial)?

7.4.4. El mal uso de los algoritmos

Jonathan Albright, profesor asistente de comunicaciones de la Universidad Elon de Carolina del Norte, investigó cómo los sitios web de extrema derecha han difundido falsas informaciones recurriendo a las debilidades de los algoritmos de búsqueda como el de Google.

Tomé una lista de estos falsos sitios de noticias que circulaban; tuve una lista inicial de 306 de ellos y usé una herramienta -como la que usa Google- para rastrearlos y luego los mapeé. Miré a donde los enlaces apuntaban - como YouTube y Facebook - y entre ellos, millones de ellos ... y simplemente no podía creer lo que estaba viendo.

"Han creado una web que está sangrando a través de nuestra web. Esto no es una conspiración. No hay una persona que haya creado esto. Es un vasto sistema de cientos de sitios diferentes que están utilizando todos los trucos que todos los sitios web utilizan. Están enviando miles de enlaces a otros sitios y juntos esto ha creado un vasto sistema satelital de noticias de la derecha y propaganda que ha rodeado completamente al sistema de medios de comunicación (en C.Cadwalladr, The Guardian, 4/12/2016).

Encontré 23.000 páginas y 1,3 millón de hipervínculos. El gráfico que sigue muestra esta nube: "

Lo que este mapa muestra es la red de distribución y se puede ver que está rodeando y realmente asfixiando el ecosistema de noticias mainstream. Cuando se mira en 3D, en realidad se ve como un virus. Y Facebook fue sólo uno de los anfitriones de este virus, que ayuda a propagar más rápidamente. (*ibidem*)

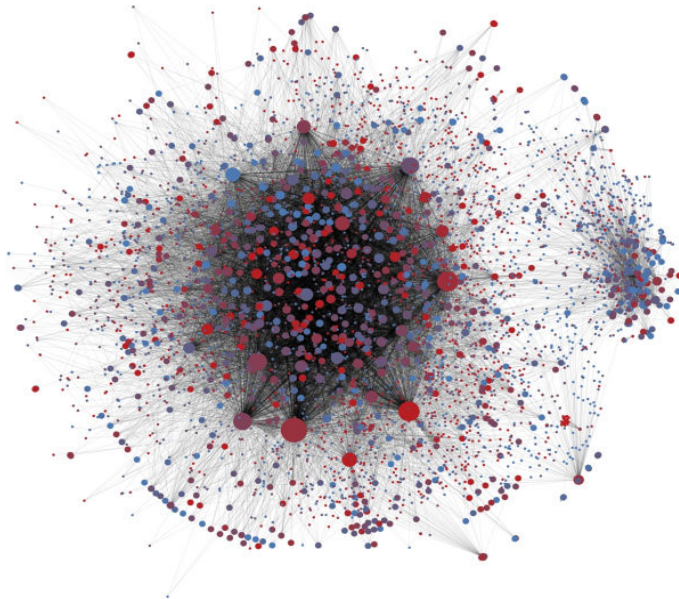


Gráfico de Albright (*The Guardian*)²⁹

Charlie Beckett, profesor de la Escuela de Medios y Comunicaciones de LSE, dice: "Lo que Jonathan Albright ha demostrado es que esto no es un subproducto de Internet, ni siquiera se hace por razones comerciales, es motivado por la ideología, por personas que tratan deliberadamente de desestabilizar Internet". Y para ello usan los mismo trucos que los sitios comerciales para engañar al algoritmo de Google: cuantas más personas buscan información sobre un tema determinado, más gente verá enlaces a los sitios correspondientes y subirá en los resultados de Google. El mapa muestra lo bien que lo están haciendo.

Cada vez que a alguien le gusta uno de estos posts en Facebook o visita uno de estos sitios web, los scripts le siguen a través de la Web. Esto permite, por la minería de datos, influir en empresas como Cambridge Analytica marcar a los individuos, seguirlos en la web y enviarles mensajes políticos altamente personalizados. Es una máquina de propaganda, dirigida a las personas individualmente para reclutarlas a una idea; es un nivel de ingeniería social que nunca había visto antes. (*ibidem*)

Si es un sitio de odio, como ocurre con los sitios neonazis, aparecerá más en la primera página de resultados, como demostró Carole Cadwalladr, la autora del artículo de *The Guardian*. Ella tipeó en Google Search "are jews" (son los judíos) y el sistema le propuso completar la frase con cuatro diferentes adjetivos. El cuarto era "evil" ("malvados") y ella ordenó esa búsqueda. Obtuvo una página entera de referencias que ofrecían un 90% de respuestas que confirmaban dicha afirmación. La primera era de una página que ofrecía los 10 mayores motivos de odio ("Top 10 Major Reasons Why People Hate Jews"), y la tercera remitía a un sitio neo-nazi. Se sorprendió descubriendo algo parecido con el tema de las mujeres ("women are..."), cuya primera propuesta para completar la pregunta fue "are evil" (son malvadas) y los 10 primeros resultados respondían afirmativamente. ¡El primero de ellos afirmaba incluso que "Toda mujer tiene algún tipo de prostituta en ella. Cada mujer tiene algo malvado en ella ... Las mujeres no aman a los hombres, les encanta lo que pueden hacer para ellas. Es razonable decir que las mujeres sienten atracción, pero no pueden amar a los hombres."!

En español no se obtienen estos mismos resultados, aunque al buscar por Hitler, una de las sugerencias que se ofrecen al usuario es "Hitler era bueno". Si se consultaba en noviembre de 2016 si ocurrió el Holocausto nazi, la primera entrada hablaba de "La farsa del holocausto", aunque a continuación

aparecía un enlace a la Wikipedia con la explicación detallada de lo que defiende esta corriente ("pantallazo" de Google a continuación).



Google ¿ocurrió el holocausto?

Todo Imágenes Vídeos Noticias Maps Más ▾ Herramientas de búsqueda

Aproximadamente 384.000 resultados (0,46 segundos)

55 Preguntas Acerca de La Farsa del Holocausto
www.bibliotecapleyades.net/sociopolitica/esp_sociopol_zion24.htm ▾
1) ¿Que prueba hay de que los nazis mataron deliberadamente a 6 millones de judíos? Ninguna. La única evidencia la constituye el testimonio de ...

Negacionismo del Holocausto - Wikipedia, la enciclopedia libre
https://es.wikipedia.org/wiki/Negacionismo_del_Holocausto ▾
El negacionismo del Holocausto es una corriente de pensamiento que tiene la finalidad de Los líderes de Hamas han promovido también la negación del Holocausto: Abdel Aziz al-Rantisi declaró que el Holocausto nunca ocurrió, que los ...
Terminología · Historia · Afirmaciones y opiniones de ...

Fuente: TICbeat, 13/12/2016

Aunque la compañía confirmó inicialmente que no tomaría medidas para que los resultados cambien ante esta pregunta (TICbeat, 13/12/2016), terminó anunciando, a fines de diciembre, que estaba cambiando su algoritmo para que ese tipo de resultado no volviera a aparecer (Wwwwhat's New, 26/12/2016).

Lo encontrado por Albright nos ha de alertar una vez más acerca del peligro de creer en la inocuidad y objetividad de los algoritmos (y en los resultados de Google Search), aunque los expertos en estadística puedan demostrar que los algoritmos predictivos son superiores a la intuición e incluso a las opiniones de expertos, pero también que los hechos negativos tienen más peso psicológico que los positivos (ver al respecto D.Kahnemen, pp. 291-305 y 491-530).


7.4.5. Viralización de información falsa y censura

En los meses previos a las elecciones en Estados Unidos hubo una oleada de noticias falsas. Según un análisis del periodista Craig Silverman, durante ese tiempo las 20 historias falsas en circulación superaron las 20 primeras historias de los 19 principales editores, "viralizadas" en las redes sociales (NewScientist, 7/12/2016).

Al principio de diciembre 2016, la Comisión Europa ha reiterado su exigencia de que Facebook, Google y Twitter vigilen con más ímpetu la incitación al odio que se produce en estas redes sociales (TICbeat, 5/12/2016). La Comisión está consciente de la viralidad de muchas opiniones odiosas y noticias falsas, como lo que ocurrió con el "#pizzagate" durante la campaña electoral norteamericana. Se difundió entonces como "noticia" que la candidata demócrata Hillary Clinton y su jefe de campaña estaban liderando una red de tráfico de niños que operaba en la pizzería Comet Ping Pong de Washington DC. Esta noticia se viralizó en las redes sociales en gran parte por el efecto "bola de nieve" de los algoritmos, causando numerosas amenazas contra el dueño de la pizzería y terminó empujando a un hombre de 28 años a entrar ahí disparando sus armas, felizmente sin herir a nadie. El portavoz de la Casa Blanca, Josh Earnest, se hizo eco de las preocupaciones que este tipo de hecho (o "realidad alterna") ha de generar: "Esto es algo sobre lo que creo que todo el mundo debería pensar... la próxima administración también va a tener que pasar algún tiempo pensando en eso" (El Mercurio, 6/12/2016).

Philadelphia Inquirer
21 mins · 🌐

"We want to stand up for white America — we're the backbone of this country, the white working-class people," she said last month.



Pistachio Girl has been fired from her Citizens Bank Park job
Pistachio Girl has been fired from her job as a food vendor at Citizens Bank Park.
PHILLY.COM

👍 🤔 🗨️ 11 5 Comments

👍 Like 💬 Comment ➦ Share

Facebook Survey ✕

To what extent do you think that this link's title uses misleading language?

Not at all Slightly Somewhat Very much Completely

Frente a una situación como la anterior, Facebook reconoce que impone censura para «evitar que la red, de entrada, se convierta en un nido de pornografía y violencia». Además, ha anunciado «un algoritmo que detecta desnudez, violencia o cualquier otra cosa que no esté de acuerdo a nuestras políticas», según un comunicado de prensa de diciembre de 2016, pero declinaron referirse sobre si este proyecto puede o no detectar noticias falsas y no está claro a partir de cuando estaría operativo (FayerWayer, 6/12/2016). También inició la prueba de una especie de encuesta que pregunta al usuario «¿Hasta qué punto crees que este artículo usa lenguaje engañoso?». Dicha encuesta aparece inmediatamente después del enlace e incluye respuestas que van desde «nada en absoluto» a «completamente», si bien el usuario puede optar por no contestar (Genbeta, 6/12/2016; ejemplo a continuación).

También han empezado a trabajar con organizaciones de verificación de datos que adoptan el código de principios de Poynter. Si una historia ha sido reportada como falsa, aparecerá un enlace al artículo explicando por qué genera disputa (@cdperiodismo, 15/12/2016; ver adjunto).

Sin embargo, según ABC.es,

A lo largo de la trayectoria de la mayor red social han sido muchas las ocasiones en las que se ha aplicado una dudosa moral a la hora de ocultar ciertos contenidos a los usuarios. ...

La pregunta es si Facebook debería aplicar otro tipo de normas morales a la hora de aplicar su criterio de selección de contenidos. «Podría aplicar las mismas normas que un medio de comunicación convencional», considera Miguel Ángel Jimeno, profesor de periodismo de la Universidad de Navarra, aunque al tratarse de una empresa privada con contenidos no periodísticos puede decidir lo que quiera. «Cuando una de esas reglas llama la atención a un sector de sus usuarios, como ha sucedido ahora, el mundo 'periodístico' se subleva. Pero creo

que solo ha sido el periodístico. A los demás usuarios, el tema parece que ni les va ni les viene», añade este experto. Frente a esa situación, y de cara a evitar que este tipo de servicios se conviertan en foros de propagación de enfoques ideológicos hostiles, la propia compañía Facebook ha puesto el foco en un programa específico para luchar contra el odio digital, una expansión de los esfuerzos de la industria tecnológica para socavar la propaganda de internet de terroristas islámicos y radicales de extrema derecha. (ABC, 16/11/2016)



Sea que se analicen los datos mediante algoritmos definidos por expertos o dejando que evolucionen mediante aprendizaje de máquina, la ética exige rendir cuenta de ello y explicar el proceso.

A pesar del potencial de aumento de la eficiencia, los algoritmos alimentados por grandes datos también pueden amplificar la discriminación estructural, producir errores que nieguen los servicios a los individuos, o incluso seducir a un electorado en una falsa sensación de seguridad. De hecho, cada vez hay más conciencia de que el público debe desconfiar de los riesgos sociales que plantea la excesiva dependencia de estos sistemas y trabajar para que sean responsables.

Incluso para técnicas como los algoritmos genéticos que evolucionan por sí solos, o algoritmos de aprendizaje de máquina donde el modelo resultante no fue elaborado a mano por una persona, los resultados son moldeados por decisiones de diseño hechas por el hombre, reglas sobre qué optimizar y elecciones sobre qué datos de entrenamiento utilizar. «El algoritmo lo hizo» no es una excusa aceptable si los sistemas algorítmicos cometen errores o tienen consecuencias indeseables.

La rendición de cuentas implica la obligación de informar y justificar la toma de decisiones algorítmicas y de mitigar cualquier impacto social negativo o daño potencial. Consideraremos la rendición de cuentas a través de la lente de cinco principios básicos: responsabilidad, explicabilidad, exactitud, auditabilidad y equidad (Diakopoulos y Friedler).

Algunas empresas se dedican ahora expresamente a identificar sitios de noticias falsas, por ejemplo Integral Ad Science, DoubleVerify, AppNexus y RocketFuel (TICbeat, 16/12/2016). Un equipo de investigadores de la Universidad de Indiana también desarrolló Hoaxy, una herramienta accesible en una plataforma web, para rastrear de forma automática la propagación de noticias falsas. El usuario solo tiene que escribir en el buscador de la plataforma el tema que le interese y el servicio le devolverá un listado con todas las noticias falsas que se han publicado sobre el asunto (EIPais.com, 23/12/2016).

Como lo expresó claramente el senador por Nueva York, Daniel Patrick Moynihan *"Todo el mundo tiene derecho a su propia opinión, pero no a sus propios hechos"*.

8. ¿Una internet inteligente?

Se ha especulado acerca de la complejidad creciente de internet y del poder combinado de los recursos de inteligencia artificial cada vez más comunes e importantes en la "nube". ¿No son todas estas conexiones parecidas a las de nuestro cerebro? ¿No deberían dar lugar a la emergencia de una inteligencia global o mente digital?

Volvamos un poco atrás. Si bien las primeras conexiones entre computadoras se diseñaron según un patrón no centralizado, las primeras comunidades de internet (como los foros y BBS) fueron definidas y organizadas desde arriba. Los miembros ingresaban donde encontraban reflejados sus intereses y emergieron líderes en forma natural, como en todo grupo humano. Pero el software no era la fuente de esta organización y los moderadores de los grupos imponían las reglas. Tampoco había una real retroalimentación, las reglas permitiendo la presencia de "merodeadores" (que leen pero no aportan

nada) y de obsesivos-compulsivos (que inundan de extensos discursos repetitivos sin escuchar a nadie), lo cual generaba anarquía y, al crecer, llevaba a una pesadilla o a un trabajo agobiante para los moderadores (Johnson: 131-136). Llevado a la web, este sistema fue la cuna del *spam* y de los *trolls* que hoy perjudican a Twitter.

Quien encontró primero una solución (aproximada) para lograr una comunidad más "inteligente" fue Rob Malda, que creó Slashdot. Después de una etapa inicial que llegó cerca de la anarquía, estableció reglas que hicieron que su comunidad se regulara en forma homeostática. En lugar de aumentar su equipo de moderadores, convirtió a todos los usuarios en potenciales moderadores: así, la tarea del control de calidad quedó en manos de la comunidad. Había retroalimentación positiva (reforzándose los temas de interés) y negativa (silenciamiento de obsesivos, *trolls*, etc.) con, además, una aleatoriedad estructurada, todo ello bajo control descentralizado e interacciones entre vecinos, como ocurre en una colonia de hormigas (Johnson: 138).

"Funciona de la manera siguiente: si se ha estado durante unas cuantas sesiones como usuario registrado, el sistema envía en algún momento una alerta informando que se ha otorgado estatus de moderador (no es tan distinto de una citación judicial que llega por correo). Como en la analogía legal, los moderadores sólo sirven para un lapso determinado, y durante ese lapso tienen el poder de calificar las contribuciones de otros usuarios en una escala de -1 a 5. Pero ese poder disminuye con el uso: cada moderador tiene una cantidad limitada de puntos que puede usar para calificar las contribuciones. Una vez distribuidos esos puntos, el puesto de moderador expira.

Esas calificaciones convergen en lo que Malda llamó *karma*: si nuestras contribuciones como usuarios son bien calificadas por los moderadores, acumulamos *karma* en el sistema, lo cual nos otorga privilegios adicionales. [...]

Había creado un tipo de moneda, un sistema de precios para ciudadanos *online*. Como aseguró que los puntos se tradujeran en privilegios especiales, les otorgó un valor. Como hizo de los poderes del moderador bienes fungibles, introdujo el concepto de escasez. Con una sola de las dos cosas, la moneda no tiene valor; con ambas combinadas se obtiene un estándar para la participación comunitaria en la formación de precios que realmente funciona. [...]

La lucidez y flexibilidad del sistema salta a la vista: visitemos el sitio de Slashdot y solicitemos ver toda la correspondencia de una conversación. Si se ha iniciado unas pocas horas antes, se encontrarán probablemente varios cientos de entradas, al menos la mitad de ellas serán de excéntricos obsesivos y *spammers*. Es el destino de cualquier sitio web lo suficientemente afortunado como para recoger miles de correos por hora.

Si restringimos el umbral de calidad a cuatro o cinco, ocurre un milagro. El volumen total cae precipitadamente; en ocasiones es un volumen de gran magnitud, pero la docena o los veinte comentarios que quedan resultan una lectura más estimulante que el contenido de ningún otro sitio tradicional, donde los colaboradores y editores cobran por sus contribuciones. Es un milagro, y no tanto porque haya material de calidad en la marea interminable de correspondencia, sino porque la comunidad ha llevado a cabo el excepcional trabajo de detectar esa calidad." (Johnson: 138-140)

Cumplió de este modo con las leyes fundamentales de la emergencia: "*el comportamiento de agentes individuales es menos importante que el sistema en su conjunto*" (Johnson: 130) con un sistema de retroalimentación basado en unas pocas reglas claras que aseguraban el equilibrio (homeóstasis) en el nivel más alto (con el máximo de filtros) aunque no en los niveles bajos (con pocos o sin filtros).

Slashdot, sin embargo, tiene un defecto: lleva a una suerte de dictadura de la mayoría: los puntos de vista mayoritarios se amplifican mientras los minoritarios terminan silenciados. Bastaría con cambiar levemente el algoritmo para corregir este defecto, eligiendo por ejemplo los moderadores cuyas colaboraciones provoquen un más amplio rango de respuestas en vez de dar prioridad al "usuario medio", pero manteniendo los filtros de calidad (Johnson: 144). También podría permitirse a los lectores elegir una u otra tendencia (como una posición en una curva de Gauss) para seguir conversaciones.

¿Que es lo fundamental en este tipo de experiencia, que nos acerca a la emergencia de la inteligencia? El contenido es evidentemente importante, como también el medio de difusión con sus características (como la interactividad y los enlaces) y el entorno, pero, por sobre todo, las reglas que han de respetar cada individuo. Estos cuatro componentes son claves para la emergencia de sistemas inteligentes, desde las colonias de hormigas hasta las neuronas del cerebro y la estructura de las

ciudades. Ya lo puso en evidencia Alan Turing, uno de los "padres" de la inteligencia artificial, en un trabajo sobre la "morfogénesis", que publicó en 1954. Usando herramientas matemáticas, demostró cómo un organismo complejo podía desarrollarse sin ninguna dirección o plan maestro, a partir de pocas instrucciones.

A mediados de los años 80, dos profesores de la UCLA, D. Jefferson y C. Taylor, diseñaron un conjunto de micro-programas que imitarían la forma en que las hormigas siguen un rastro, el que llamaron "Tracker". No era una solución sino un conjunto aleatorio de posibles desplazamientos, con un sistema de retroalimentación y autocorrección que "premiaba" los mejores resultados, haciendo emerger así programas cada vez más competentes. El videojuego SimCity, *best-seller* de los años 90, de Will Wright, fue construido sobre la misma base, llevando la idea del desarrollo "*bottom-up*" fuera de los laboratorios y ayudando a popularizarla.

La clave de las escasas instrucciones que conforman las reglas de comportamiento de cada individuo es la detección de patrones en el entorno. En SimCity (vista de la versión original al lado), cada célula se conecta con las vecinas y ajusta su comportamiento en respuesta a la conducta de estas.



"Los algoritmos en sí son relativamente simples — fíjate en el estado de tu vecino y actúa en consecuencia—, pero la magia de la simulación es posible porque la computadora hace miles de cálculos por segundo. Dado que cada célula influye en el comportamiento de otras, los cambios parecen transmitirse por todo el sistema con una fluidez y una definición que sólo pueden describirse como «propias de la vida».

Es impactante el parecido con nuestras hormigas y embriones. Cada manzana de edificios en SimCity obedece a un rígido conjunto de instrucciones que marca su conducta, del mismo modo que nuestras células consultan la hoja de ruta de nuestros genes. Pero esas instrucciones dependen de las señales recibidas desde otros bloques en el vecindario, al igual que las células «espían» a través de las uniones celulares para captar el estado de sus vecinas. Con sólo unos pocos bloques, el juego es mortalmente aburrido y parece un robot poco convincente. Pero con miles de bloques, que responden a docenas de variables el paisaje de la ciudad simulada cobra vida: surgen barrios ricos y deprimidos en función de recesiones virtuales o de auges repentinos. Lo mismo que sucede con las colonias de hormigas, aunque es diferente." (Johnson: 80)

El análisis de *big data* es una de las consecuencias de este desarrollo científico: ha heredado esta idea y busca detectar patrones demográficos para regular comunidades *online* y orientar el marketing (Johnson: 60). Pero está muy lejos de acercarse a la emergencia de la inteligencia aunque esté usando *software* de la llamada inteligencia artificial.

¿Se parece de algún modo internet a una estructura como la de una colonia de hormigas o como SimCity? ¡Ni de lejos! Es tan solo "*una vasta suma de datos interconectados*" (Johnson: 102).

"En la medida en que la Web conecte a seres más conscientes que cualquier otra tecnología anterior, se la podrá considerar como un tipo de cerebro global. Pero tanto los cerebros como las ciudades hacen algo más que conectar, porque la inteligencia requiere tanto conectividad como organización. En el mundo real, muchos sistemas descentralizados generan espontáneamente una estructura cuando incrementan su tamaño: las ciudades se organizan en barrios o en ciudades satélites; las conexiones neuronales de nuestros cerebros desarrollan regiones extraordinariamente especializadas. ¿Durante los últimos años, la Web ha recorrido un camino de desarrollo comparable? ¿Está volviéndose más organizada a medida que crece? [...]

Imaginemos el universo de documentos HTML como una ciudad que se extiende a través de un vasto paisaje, donde cada documento representa un edificio en el espacio. La ciudad de la Web sería más anárquica que cualquier ciudad del mundo real en el planeta: sin referencias de comercios relacionados entre sí o de negocios afines; sin barrios de teatros o de carnicerías; sin comunidades bohemias o típicas casas de lujo; sin ni siquiera los desafortunados barrios marginales de la *edge city* de Los Angeles o de Tyson's Corner. La ciudad de la Web sería simplemente una masa de datos indiferenciada que se hace más confusa con cada nuevo

«edificio») que se levanta. Y si es cierto que la Web sería una ciudad lamentable, mucho peor sería como cerebro. [...]

El cerebro requiere de grupos específicos para dar sentido al mundo, y esos grupos sólo emergen de un complejo intercambio entre las neuronas, el mundo exterior y nuestros genes (por no mencionar unos miles de factores más). Algunos sistemas, tales como la Web, son genios de la conexión pero pésimos con la estructura. Las tecnologías que sustentan Internet —desde los microprocesadores en cada servidor de la Web hasta los protocolos de final abierto que gobiernan los propios datos— han sido brillantemente diseñadas para manejar grandes incrementos de escala, pero son indiferentes, si no lisa y llanamente hostiles, a la tarea de crear un orden de nivel superior.” (Johnson,: 105-106)

¡Necesitamos a Google para no perdernos, pero Google no introduce ninguna organización! Existe sin embargo una experiencia que podría acercarnos más a lo que podría ser una red inteligente: Alexa, actualmente una subsidiaria de Amazon. Fue creada por Brewster Kahle y Bruce Gilliat en 1996. Es un buscador que sigue su usuario mientras navega entre las páginas que encontró y registra los enlaces que siguió. Así,

El software aprende por observación de la conducta de los usuarios de Alexa: si cien usuarios visitan Feed y luego van a Salón, entonces el *software* comienza a percibir una conexión entre los dos sitios, una conexión que puede debilitarse o fortalecerse de acuerdo con el rastreo de la conducta. En otras palabras, las asociaciones no son el trabajo de una conciencia única sino la suma total de miles y miles de decisiones individuales, una guía en la Web que se crea siguiendo el rastro de un número inimaginable de huellas. (Johnson: 110-111)

El poder aparente de agrupación de Alexa emerge de la navegación de los usuarios, ninguno de los cuales busca crear grupos o categorías de sitios con elementos en común. Usan el sistema y el sistema “aprende” de ellos, haciéndose más organizado, más “inteligente” a medida que más personas lo usan. Mil usuarios no generan nada, pero si son más de diez mil, empiezan a aparecer asociaciones con real sentido. Es obviamente una forma de inteligencia muy diferente de la que nos proporciona nuestro cerebro. Alexa no comprende nada; no lee ni estudia el diseño de los sitios web: solo busca patrones en los enlaces seguidos... y esto puede tener resultados inesperados, como le ocurrió a un niño que pedía – por su asistente de voz – una canción infantil y, al pronunciar mal, obtuvo una advertencia de pornografía (FayerWayer, 30/12/2016). Su inteligencia es “*la sabiduría agregada de las miles o millones de personas que usan el sistema*” (Johnson: 111) pero no se parece en nada a ellas.

Un sistema de información adaptador, capaz de reconocer patrones complejos, podría ser uno de los inventos más importantes de la historia de la humanidad. ¿Qué importancia tiene que no aprenda jamás a pensar por sí mismo?

Un programa de *software* emergente que rastrea asociaciones entre sitios Web o discos compactos de audio, no escucha música; registra patrones de compra o hábitos de consumo que nosotros ofrecemos y nos permite interactuar con la guitarra y el canto. En un nivel básico de lo humano, parece ser una diferencia que vale la pena preservar y que quizá nunca podamos trascender ni siquiera dentro de cien años o más. (Johnson: 115)

Como ya dicho en el Capítulo 4, falta muchísimo para que llegemos a la “singularidad” que pronostica Ray Kurzweil, donde las máquinas inteligentes nos superen (y ni siquiera sabemos si es posible).

Conclusión

La producción y difusión del conocimiento es hoy tan abundante que es prácticamente imposible seguirlo aprovecharlo sin nuevas herramientas. Pero los algoritmos no parecen ser el mejor instrumento para extraer lo relevante. Operan en forma automática, como lo hace también el pensamiento humano de tipo intuitivo, pero no solo en base a su propia experiencia, sino en base a los sesgos que introducen sus programadores y otros usuarios (no siempre bien intencionados):

Los algoritmos no existen en el vacío. Requieren seres humanos. [...]

Todo en Internet es una mezcla de lo humano e inhumano. Los algoritmos automatizados desempeñan un papel muy grande en algunos servicios como, por ejemplo, el motor de búsqueda de Google. Pero los humanos juegan un papel en estos servicios también. Los seres humanos operan la lista blanca y la lista negra de sitios en el motor de búsqueda de Google. Hacen lo que podría pensarse como decisiones manuales, en parte debido a que los algoritmos de hoy en día son tan defectuosos. Lo que es más – y esto afirma lo que debería ser obvio –: los seres humanos escriben los algoritmos. Eso no es insignificante. Lo que significa es que los algoritmos conllevan sesgos humanos. Llevan los prejuicios de las personas que los escriben y de las empresas por las que trabajan esas personas (Wired, 2/09/2016).

Es posible que esto mejore con la inteligencia artificial basada en redes neuronales artificiales, que imita la forma en que opera nuestro propio cerebro para aprender (autoponderando sus conexiones en función de las experiencias repetidas). Pero convendrá sin duda mantener siempre un enfoque crítico y explorar otras posibles alternativas si es que surgen.

La misma experiencia del autor (descrita al final del capítulo 2) ha puesto en evidencia que el trabajo de análisis de datos, cualquiera sea el método utilizado, hereda múltiples sesgos:

1. Elección de fuentes: nadie puede acceder a todas las fuentes y todos tenemos que realizar algunas elecciones antes de empezar a recopilar datos; lo cual es válido en cualquier ámbito.
2. Las fuentes mismas ya hicieron su propia selección de los datos que han recopilado y difundido.
3. Para las fuentes que consideramos (o consideran que quieren aplicar) las técnicas de análisis, elegimos informaciones y descartamos otras, de acuerdo con nuestros intereses, criterios, conocimientos previos y creencias. Ni los mejores científicos están exentos de ello, siempre influidos por su concepción del mundo, que la expliciten o no (de ahí una rama de la ciencia que se llama “investigación de fundamentos”).
4. Finalmente, realizamos el análisis e interpretamos los resultados de acuerdo a este mismo marco mental.

Otro aspecto clave en la situación actual es el de la amplia “observación” de los usuarios por parte de quienes prestan sus servicios, una situación de “vigilancia generalizada”, como denunciaba ya Rheingold en 2002:

La vigilancia generalizada suscita varias cuestiones importantes:

¿Quién espía a quién? ¿Quién tiene derecho a conocer esa información?

¿Quién controla la tecnología y sus usos? ¿El usuario, el gobierno, el fabricante, la compañía telefónica?

¿Qué tipo de personas seremos cuando utilicemos la tecnología? (112)

Por esta razón los proyectos de redes distribuidas son de la mayor importancia, ya que pueden conducir no solo a la liberación frente a dicha vigilancia (que podría crecer exponencialmente con la “internet de las cosas”) sino, a mediano plazo, a una organización social diferente:

¿El usuario podrá utilizar las tecnologías de las multitudes inteligentes para saber todo lo que necesita sobre el mundo en el que vive y para conectarse con los grupos que pueden beneficiarle? ¿Le permitirán cooperar con cualquier persona que elija con ayuda del ordenador que lleva puesto? ¿O podrán los demás averiguar todo lo que necesitan sobre el usuario en cuestión, a través de los sensores con los que éste tropieza y la información que emite? Las diversas respuestas a estas preguntas conducen a futuros diferentes. (Rheingold:113)

Hemos visto que existen formas alternativas de personalización, sin ingresar los datos personales en enormes bases de datos: Apple ha mostrado una posible modalidad.

Como planteó Steve Mann (aunque pretende combatir la tecnología con tecnología), el objetivo final no puede ser sino “mejorar la inteligencia de la especie”, dar a cada ser humano un instrumento para ser más inteligente, lo cual obliga a pensar la tecnología y su línea de desarrollo antes de seguir en forma irreflexiva el impulso de lo ya hecho. Como dice Joshua Cooper, “los tecnólogos no saben de humanidad” (222), y ésto es algo que debería cambiar. ¡Luchemos por ello, especialmente en la educación superior que los forma y en las políticas que los pueden orientar!

Bibliografía

- Alba, D. (2016): *Repeat After Me: Humans Run the Internet, Not Algorithms*, Wired, 2/09/2016, <http://www.wired.com/2016/09/facebook-trending-humans-versus-algorithms/>
- Asigra: *What is Big Data* (infografía) <https://cblconsultics.files.wordpress.com/2015/02/big-data-infografia.jpg>
- BBC: "15 key moments in the story of artificial intelligence", <http://www.bbc.co.uk/timelines/zq376fr>
- Biewald, L.: *How machine learning will affect your business*, ComputerWorld, 20/10/2015, <https://goo.gl/Z7NrY0>
- Cadwalladr, C. (2016): *Google, democracy and the truth about internet search*, The Guardian, 4/12/2016. <https://goo.gl/WAejrC>
- Camarer-Cano, L. (2015): *Comunidades tecnosociales. Evolución de la comunicación analógica hacia la interacción analógicodigital*, Revista Mediterránea de Comunicación, vol.6(1), 187-195.
- Catlin, J.: *Text Analytics: The Next Generation of Big Data*, Inside Big data, 5/06/2015, <https://goo.gl/6d5hFt>
- Colle, R. (2015): *Secuestradores y secuestrados: Nuestros datos en internet*, auto-edición, en ISSUU. <https://goo.gl/VXEt5w>
- Colle, R. (2014): *Internet ayer, hoy y mañana*, auto-edición, en ISSUU: <https://goo.gl/wiSLbm>
- Colle, R. (2013): *Prensa y Big Data: El desafío de la acumulación y análisis de datos*, Revista Mediterránea de Comunicación, 4(1).
- Colle, R. (2002): *Explotar la información noticiosa - "Data Mining" en el Periodismo*, Depto de Biblioteconomía y Documentación, Universidad Complutense de Madrid, 2002. Disponible en ISSUU.
- Collins, H.: *Experts artificiels - Machines intelligentes et savoir social*, Paris, Seuil, 1992.
- Cooper, J. (2016): *The Seventh Sense: Power, Fortune, and Survival in the Age of Networks*, Boston, Little, Brown & Co.
- De Maeyer, J. (2016): *Objectivity revisited*, Nieman lab, 20/12/2016. <http://www.niemanlab.org/2015/12/objectivity-revisited/>
- Diakopoulos, N. & Friedler, S. (2016): *How to Hold Algorithms Accountable*, MIT Technology Review, 17/11/2016) <https://www.technologyreview.com/s/602933/how-to-hold-algorithms-accountable/>
- Fauscette, M. (2014): *Transforming Data Into Action*, <http://www.mfauscette.com/.../transforming-data-into-action-part-one.html>
- Freedman, D. (1995): *Los hacedores de cerebros*, Santiago de Chile, A.Bello.
- Hammond, K. (2015): *Why artificial intelligence is succeeding: Then and now*, ComputerWorld, 14/09/2015.
- Greenslade, R. (2016): *Why Facebook is public enemy number one for newspapers and journalism*, The Guardian, 20/09/2016. <https://goo.gl/dASWJz>
- Jiménez de Luís, A. (2014): *Internet nos ha convertido en la generación transparente*, El Mundo.es, <http://www.elmundo.es/tecnologia/2014/11/20/546dfb>
- Kahneman, D. (2012): *Pensar rápido, pensar despacio*, Barcelona, Debate.
- Knight, W.: *Watsoncito, Watsoncito, ¿hacia dónde debe IBM redirigir su negocio?*, MIT Technology Review, 7/11/2014.
- Kurzweil, R. (2012): *La singularidad está cerca*, Lola Books (Original: *The Singularity is Near*, Viking Press, 2005)
- Lee, R. (2014): *Privacy, big data and analytics: A perfect storm*, IBM Big Data & Analytics Hub, 6/06/2014, descargado el 18/06/2014 de <http://www.ibmbigdatahub.com/blog/privacy-big-data-and-analytics-perfect-storm>

- López-Cantos, F.: *La investigación en comunicación con metodología BigData*, Revista Latina de Comunicación Social, 70, pp.878 a 890. <http://revistalatinacs.org/070/paper/1076/46es.html>
- Mann, S., & Niedzviecki, H. (2001): *Cyborg: Digital Destiny and Human Possibility in the Age of the Wearable Computer*, Toronto, Doubleday Canada.
- Rainie, L. & Wellman, B. (2012): *Networked: The New Social Operating System*, MIT Press, Cambridge (MA).
- Rheingold, H. (2004): *Multitudes inteligentes*, Barcelona, Gedisa.
- Rodríguez, A.: *El Big Data ya no es solo cosa de los grandes*, <http://hipertextual.com/2015/11/startups-big-data> Hipertextual, 20/11/2015.
- Pastor, J.: *La Skynet benévola: los ordenadores "conscientes" están cada vez más cerca*, Xataka, 9/12/2014.
- Penalva, J.: *Este es uno de los nuevos exámenes que tendrán que aprobar las máquinas avanzadas*, Xataka, 11/03/2015.
- Penrose, R. (2007): *Las sombras de la mente*, Barcelona, Crítica, 2007.
- Pratt, I. (1994): *Artificial Intelligence*, Londres, Macmillan.
- Sadowski, J. (2016): *Companies are making money from our personal data – but at what cost?*, The Guardian, 31/08/2016. <https://goo.gl/dnbAMF>
- Sánchez-Migallón, S. (2015): *El gran debate sobre si será posible o no una inteligencia artificial*, Xataka, 18/08/2015.
- Schank, R. (1986): *El ordenador inteligente*, Barcelona, A.Bosch.
- Searle, J. (1985): *Mentes, cerebros y ciencia*, Madrid, Cátedra.
- Simonite, T. (2014): *Demis Hassabis, Founder of DeepMind Technologies and Artificial-Intelligence Wunderkind at Google, Wants Machines to Think Like Us*, MIT Technology Review, 3/12/2014.
- Schmidt, E. y Cohen, J. (2014): *El futuro digital*, Madrid, Anaya.
- Sey, A. & Castells, M. (2004): *From Media Politics to Networked Politics: The Internet and the Political Process*, en Castells, M. (ed.): *The network society: A cross-cultural perspective*, Cheltenham, Reino Unido; Northampton, MA: EdwardElgar Pub.
- Stucke, M.E. & Ezrachi, A. (2016): *The Subtle Ways Your Digital Assistant Might Manipulate You*, Wired, 29/11/2016. <https://www.wired.com/2016/11/subtle-ways-digital-assistant-might-manipulate/>
- Tierney, J. (2014): *Customer Data Privacy has Become an Everyman Problem*, Loyalty36org, 14/03/2014, <http://loyalty36.org/resources/article/customer-data-privacy-has-become-an-everyman-problem>
- Trejo, R. (2003): *La Internet como bien público*, Revista Diálogos de la Comunicación, Nº. 67, pp.29-40. <http://dialogosfelafacs.net/wp-content/uploads/2012/01/67-revista-dialogos-la-internet-como-bien-publico.pdf>
- UIT (2014b): *Informe sobre Medición de la Sociedad de la Información 2014 - Resumen Ejecutivo*, Ginebra, UIT, <https://goo.gl/kO4Ggq>
- Westfal, Ch. & Blaxton, T.: *Data mining solutions*, New York, Wiley, 1998.
- Wright, R. (2006): *Nadie pierde, La teoría de juegos y la lógica del destino humano*, Barcelona, Tusquets.
- Zamorano, E.: *Conoce las lúbricas imágenes que crea una inteligencia artificial tratando de reconocer fotos*, FayerWayer, 19/06/2015.
- Zanoni, L. (2014): *Futuro inteligente*, Autoedición, descargado el 13/01/2015 de <http://www.futurointeligente.com.ar/> (Disponible en varios formatos)

1. Un gigabyte (GB) equivale a 10^9 bytes, un terabyte (TB) equivale a 10^{12} bytes o 1.000 GB, un petabyte (PB) a 10^{15} bytes o 1.000 terabytes y un exabyte (EB) equivale a 10^{18} bytes.

2. Vea mi libro anterior "Secuestradores y secuestrados: Nuestros datos en internet" (disponible en ISSUU).

-
3. He propuesto un model analítico sencillo de este tipo para constituir bases de datos periódísticas y lo he aplicado por años. Véase mi libro "Explotar la información noticiosa" y mi artículo "Análisis lógico de hechos noticiosos", Revista Latina de Comunicación Social, año 2000, n.27 (<http://www.revistalatinacs.org/ada2000tma/126colle.html>)
 4. "DBA – Security Superhero: 2014 IOUG Enterprise Data Security Survey"
 5. <http://data.bsa.org/#section-press>
 6. Empresa que provee este tipo de servicio de análisis.
 7. Publicado por el Depto. De Biblioteconomía y Documentaciónm Fac. De Ciencias de la Información, Universidad Complutense de Madrid (Ver Bibliografía para su descarga en PDF).
 8. Realizado con el software GVA: *Generic Visualization Architecture*.
 9. "5.000 noticias de tecnologías digitales de comunicación", Razón y Palabra, dic.2002-enero 2003.
 10. Se alude al formato necesario para la inserción en una base de datos donde se registra en diferentes "campos de datos" las variables que se analizarán.
 11. *Artificial intelligence is hard to see - Why we urgently need to measure AI's societal impacts*", Medium.com, 11/09/2016.
 12. Alusión a críticas ocurridas antes de la campaña electoral por la presidencia de Estados Unidos. Ver "Repeat After Me: Humans Run the Internet, Not Algorithms", <https://goo.gl/VfNORU>
 13. Los nacidos entre los años 1981 y 2000.
 14. <https://www.humanbrainproject.eu/>
 15. "Internet of Things; Requires Big Data to be turned upside down to become Smart Data", nov.2014
 16. Yekhanin, S. & Dwork, C.: Database Privacy, <https://goo.gl/akpb55>
 17. Este apartado combina parte de mi artículo de la Revista Mediterránea de Comunicación de 2013 y de un capítulo de mi libro "Sistémica de los medios de comunicación en la era de las redes" (ver Bibliografía).
 18. Jaime Irazo, Javier M. Buldú and Jacobo Aguirre: *Competition among networks highlights the power of the weak*. Nature Communications, 2016
 19. Fórmula propia de la teoría de juego, que es una de las alternativas posibles: te doy según lo que me diste (retribución).
 20. "El teléfono móvil y la multitud: política mesiánica en la historia filipina reciente"
 21. *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*", Little, Brown.
 22. Los juegos de suma cero son los donde la ganancia o pérdida de un participante se equilibra con exactitud con las pérdidas o ganancias de los otros participantes. La suma no cero es donde todos ganan o se gana más de lo que eventualmente se pierde.
 23. En J.P.Changeux y J.Chavaillon: *Origins of the human brain*, Oxford, Clarendon Press, 1995
 24. Cfr. su libro "Sistemas emergentes" (ver bibliografía).
 25. Página personal: <http://wearcam.org/bio.htm>
 26. http://futureoflife.org/misc/ai_conference
 27. <https://goo.gl/CuNVNf>
 28. <http://www.inc.com/brian-oconnor/how-wearables-will-make-your-brain-superior.html>
 29. A plena pantalla se pueden observar los nombres de los sitios de medios de prensa americanos en la parte central.



Colección Mundo Digital

de Revista Mediterránea de Comunicación

