

EL PAÍS

ARCHIVO

EDICIÓN  
IMPRESA

MIÉRCOLES, 3 de junio de 2009

REPORTAJE:

## Tus datos íntimos son una mina

La digitalización de la vida cotidiana genera miles de registros - La 'minería de datos' sirve para la investigación médica o la lucha contra el terror - Las cifras anticipan nuestros actos

ABEL GRAU | 3 JUN 2009

**Archivado en:** AEPD Ley protección datos Confidencialidad Terrorismo internacional Investigación médica IBM Tecnología Google Comercio electrónico Política antiterrorista Internet Empleo Lucha antiterrorista Telecomunicaciones Empresas Investigación científica Informática Industria Economía Trabajo

En nuestro quehacer cotidiano, cada vez más digitalizado, generamos constantemente datos sobre nuestros hábitos. Dejamos pistas en todas partes. Al comprar en el supermercado pasamos la tarjeta de fidelidad y, clic, quedamos registrados: cada semana, por ejemplo, junto a la carne, la verdura y los huevos, solemos comprar una cuña de parmesano y, sólo una vez al mes, un cartón de helado de vainilla. Luego enviamos SMS con el móvil y, clic, nuestro registro indica con cuánta gente nos relacionamos y si somos un usuario frecuente. Quizá después, en la oficina, navegamos un rato; leemos un par de periódicos, mandamos tres *e-mails* y compramos un billete de tren; clic, clic, clic...

**Habrán modelos nuestros como amantes, pacientes y empleados**

**Google puede conocer ya el estado de ánimo de su plantilla**

Las bases de datos aumentan masivamente hasta convertirse en un pandemio de millones de entradas. De ese caos, sin embargo, es posible desentrañar pautas sobre nuestro comportamiento para elaborar modelos predictivos y aplicarlos a la publicidad y al *marketing*, claro, pero también a la medicina, la gestión empresarial, los procesos electorales y la lucha contra el terrorismo. A la búsqueda de esas pautas se dedican los expertos en *data mining* (explotación o *minería de datos*), básicamente especialistas en estadística, matemáticos e ingenieros informáticos, encargados de desbrozar semejantes junglas de archivos en busca de patrones de comportamiento que ayuden a anticipar nuestras decisiones.

"El floreciente mundo de los datos es un laboratorio gigante sobre el comportamiento humano. Es una mesa de operaciones para las ciencias sociales, el comportamiento económico y la psicología", escribe Stephen Baker, periodista de *BusinessWeek*, en el ensayo *The Numerati. How they'll get my number and yours* (Jonathan Cape). En su libro, Baker aborda el prometedor futuro de la *minería de datos*: "En la década que viene, cada uno de nosotros dará lugar, a menudo, sin advertirlo, a modelos propios en casi cada aspecto de la vida. Seremos modelados como trabajadores, pacientes, soldados, amantes, clientes y votantes", escribe. Su vaticinio es ambicioso: sumadas todas estas facetas, veremos la "modelación matemática de la humanidad".

Una empresa habituada a trabajar con enormes bases de datos, el buscador Google, ya está aplicando el *data mining* en la gestión de los recursos humanos de su propia plantilla. Ante la marcha de algunos cargos estratégicos (ingenieros, diseñadores y ejecutivos de ventas), la compañía está desarrollando una herramienta que le permitiría conocer el estado de ánimo de sus trabajadores y averiguar cuáles se plantean marcharse, según publicó recientemente *The Wall Street Journal*. Consiste en un algoritmo (una fórmula matemática) elaborado a partir de un gran volumen de datos (entrevistas, registros de nóminas y ascensos y evaluaciones de compañeros, etcétera) de sus 20.000 trabajadores. Google no da muchos detalles más, pero

el funcionamiento de una herramienta así es conocido. "Se trata de crear un modelo de predicción del comportamiento", señala Lluís Belanche, profesor titular del Departamento de Lenguajes y Sistemas Informáticos de la Universidad Politécnica de Cataluña.

"Se toma como referencia a los empleados que se fueron de la empresa porque no se sentían valorados. Cuando se tiene una muestra amplia, de miles, se analiza cómo se comportaron poco antes de marcharse". ¿Quizá mandaban menos *e-mails* a tan sólo un reducido grupo de amigos? ¿Pasaban más tiempo visitando páginas *web* o leyendo la sección de deportes?, se pregunta Baker, desde Nueva Jersey. "Si existe alguna pauta común entre estas personas insatisfechas, puedes enseñar a una máquina a buscar entre las pautas de la plantilla actual, y localizar a los empleados que encajan en ese perfil". Se trata de proyecciones y no garantizan una fiabilidad del cien por cien, advierte Belanche, "pero cuantos más registros se tienen, más confianza merece la predicción".

Las técnicas no son nuevas, pero por primera vez se empiezan a aplicar a gran escala a la gestión de los recursos humanos, la investigación médica y la lucha contra el terrorismo. "El *data mining* de Google no es tan diferente de otros análisis que hacen las empresas en sus intentos de anticipar nuestro comportamiento", señala Baker. "Las compañías de móviles, por ejemplo, examinan los patrones de nuestras llamadas en un intento de predecir quién es más probable que se cambie a otra operadora", añade. "Si tus llamadas siguen ese patrón es muy probable que recibas una oferta especial tentadora para que renueves tu contrato".

Las herramientas que procesan tal volumen de datos "son algoritmos de aprendizaje que proceden del campo de la inteligencia artificial", sostiene Fernando Corbacho, socio de operaciones de la compañía CognoData, especializada en la aplicación de la *minería de datos* en el sector empresarial. "Analizan las bases de datos y aprenden a distinguir entre los clientes que se han ido y los que no. Elaboran un perfil del cliente que se ha ido y luego lo aplica a cada empleado y dice si encaja en el perfil de fuga".

Otros gigantes del sector informático ya estudian cómo optimizar el rendimiento de sus trabajadores. La multinacional IBM se ha fijado en sus especialistas en tecnología. El proyecto consiste en reunir exhaustivos inventarios de las habilidades de cada uno de ellos y luego calcular matemáticamente la mejor manera de utilizarlos, según escribe Baker. Uno de los cerebros de la idea es el matemático Samer Takriti, investigador del Thomas J. Watson Research Center de IBM, en Nueva York, que se ha rodeado de un equipo de especialistas, desde *data miners* y estadísticos hasta psicólogos y antropólogos.

La empresa tiene vedados los archivos de personal, con evaluaciones anuales, pero prácticamente todo el resto de información está a su disposición. Y necesitan mucha. Cuanta más, mejor. Escudriñando currículos y registros de proyectos, el equipo puede construir un perfil de las habilidades y experiencia de cada trabajador. Puede saber cómo usan su tiempo y con quién se reúnen a partir de los calendarios *online*.

Además, en los registros de llamadas y los correos electrónicos están grabadas las redes sociales de cada técnico. "¿A quién envían copias adjuntas de sus *e-mails*? ¿Mandan copias ocultas a alguien? Estos mensajes ocultos podrían apuntar al crecimiento de redes informales dentro de la compañía", incide Baker. "Podrían mostrar que un cargo medio lidera en silencio a un importante equipo de colegas y que su jefe está fuera del circuito. Quizá ambos deberían intercambiar sus puestos". Proyectos como los de Google e IBM están aún en fase inicial, pero dejan entrever su alcance potencial.

En Estados Unidos la privacidad de los empleados está claramente definida: "El empresario es propietario de cada golpe de tecla y cada clic del ratón y de cada llamada del móvil de la empresa", subraya Baker. "Además pueden estudiar calendarios electrónicos, registros laborales y currículos". En España, una sentencia del Tribunal Supremo de 2007 concluyó que los empresarios pueden rastrear toda esa información pero tienen la obligación de avisar antes a los empleados de cualquier supervisión de sus comunicaciones a través de Internet, y deben establecer protocolos.

"Fue una sentencia algo ecléctica", señala el director de la Agencia Española de Protección de Datos (AEPD), Artemi Rallo. "Se proclamó que el tipo de servicios que se ponen a disposición de los empleados son dignos de protección, pero en la misma sentencia se recoge que si el empresario avisa de que podrá acceder o accederá a *e-mails* e historiales y se protocoliza, la práctica sería legal". En cualquier caso -añade Rallo-, son situaciones nuevas, sobre las que existe escasa jurisprudencia, y la cobertura legal va avanzando para adaptarse.

Los nuevos expertos en gigantescas bases de datos, los *numerati* -como los ha bautizado Baker- juegan ya una función crucial en las empresas. "Por múltiples razones", observa Javier Arias, ingeniero de alianzas estratégicas de Google en España. "Para algunas es una herramienta para encontrar nuevas fuentes de beneficios (¿qué producto tengo que desarrollar/crear), para otras, el *data mining* es parte de su investigación y desarrollo (¿qué moléculas tienen más probabilidades de funcionar contra un cáncer específico?). Y añade que las entidades bancarias pueden emplear la técnica para "detectar que una operación con tarjeta de crédito es fraudulenta porque no se corresponde con el patrón de uso normal".

Los usos comerciales del *data mining* son abundantes. "Los supermercados y las grandes cadenas, como Wal-Mart o Ikea, crean grandes bases de datos en los que cada entrada es, por ejemplo, un tique", indica Belanche. "Sabemos qué has comprado y, si tienes tarjeta de fidelidad, incluso tienen tus datos personales". Es información fundamental para determinar los patrones de compra y poder estimular las ventas. "Localizan por ejemplo dos productos que siempre se compran juntos y los colocan en dos pasillos separados para que tengas que pasar por un tercero y compres otros que no tenías previsto adquirir".

En la elaboración de los modelos de predicción se emplean desde la estadística tradicional hasta las reglas de asociación ("si este producto se compra junto a un segundo y un tercero, entonces es probable que se compre este cuarto"), indica Belanche. "También los bancos suelen invertir en el desarrollo de sistemas de predicción de impagos de préstamos", prosigue. "Las líneas aéreas en Internet cuentan con un registro de millones de entradas. Si se analiza el registro de clics se puede saber, por ejemplo, en qué momento un cliente ha dejado de comprar un billete y por qué; quizá porque ha recibido un cargo sorpresivo, por ejemplo. A las empresas les interesa saberlo".

E incluso se están estudiando aplicaciones en los procesos electorales: conocer el patrón de voto de los electores para poder personalizar el mensaje del candidato. En su ensayo, Baker describe el proyecto de la empresa Spotlight Analysis, dirigida por Josh Gotbaum, ex funcionario de las Administraciones de Carter y Clinton. Se propone localizar a los decisivos *swing voters* (el voto indeciso que puede resultar determinante en unos comicios) para poder afinar los mensajes electorales. ¿La técnica? En primer lugar, toman las colosales bases de datos de empresas como ChoicePoint o Acxiom, que acumulan registros sobre sentencias judiciales, transacciones inmobiliarias y de impuestos, y sobre hábitos de consumo (suscripciones a revistas, compra de libros, viajes, etcétera). Son compañías que "compran cada pequeño dato sobre nosotros que está en venta", advierte Baker, "y luego venden selecciones de esos datos a cualquiera que quiera perfilarnos para dirigirnos una campaña". Luego se cruzan esos registros con los análisis sobre las preocupaciones de los consumidores, suministrados por empresas como Yankelovich (que recoge lo que piensan 175 millones de consumidores sobre el éxito profesional, la vida familiar o su comunidad, entre otros valores). A todo ello se añade una serie de encuestas telefónicas. El resultado permite segmentar a los votantes de una manera muy útil para el diseño de las campañas electorales. "El tipo de análisis estadístico que discutimos aquí", precisa Baker, "tanto si se trata de predecir nuestro comportamiento como cazadores de casas o como consumidores de vino, es por naturaleza aproximado. Se basa en la probabilidad". Logran su objetivo "si consiguen respuestas mejores, más rápidas o más baratas que las habituales".

El uso de la *minería de datos* en medicina es incipiente pero prometedor. Baker enumera algunos proyectos de instalación de sensores en el domicilio de personas de edad avanzada. Recogerán las pautas de comportamiento y avisarán de cualquier irregularidad, como un descenso de peso o un cambio del ritmo de los pasos. Otros estudios investigarán durante las

próximas décadas el desarrollo de nanosensores instalados en el organismo para examinar la sangre y poder alertar a tiempo del crecimiento de un tumor. "Para que funcionen, los *numerati* tendrán que desarrollar normas estadísticas para cientos de nuestras pautas biológicas".

Con todo, el campo de la *minería de datos* que plantea mayores conflictos sobre la invasión de la privacidad es el de la lucha contra el terrorismo. Tras los atentados del 11-S de 2001, Washington empleó más de mil millones de dólares para unir las ingentes bases de datos de la CIA y el FBI y ofrecer así una única fuente a los *data miners*, según explica Baker. Y, además, -enumera el periodista- rastrearon datos de consumo y demografía, revisaron registros de líneas aéreas y recibos de hotel, junto a millones de vídeos y horas de llamadas telefónicas y tráfico de internet recopilados por la National Security Agency (NSA). E incluso, según publicaron varios medios en 2006, las grandes compañías telefónicas entregaron al Gobierno el registro de millones de llamadas.

En esa batalla contra el terror se incluyó el *software* NORA, un sistema para identificar relaciones no evidentes entre tramposos de casinos, según detalla Baker. El programa fue ideado para rastrear las bases de datos internas de los casinos en busca de vínculos entre nombres, direcciones y números de teléfonos.

Tras el 11-S se utilizó NORA para cribar las identidades y movimientos de terroristas. Su creador, Jeff Jonas, es un científico de IBM que aboga por la defensa de los derechos de privacidad. "Necesitaremos tecnología que proteja nuestra identidad y políticas que salvaguarden nuestros derechos", advierte en *The Numerati*. "Vamos a necesitar mucha gente inteligente en política". Sin una profunda regulación -añade Baker- es probable que recibamos lo peor de ambos mundos, una sociedad de la vigilancia que aun así es incapaz de mantenernos seguros.

En cualquier caso, Baker pronostica un inmenso nuevo territorio para la ciencia. "Promete ser uno de los grandes logros del siglo XXI. Su alcance crecerá hasta incluir la mayoría del mundo físico a medida que los matemáticos metan mano en nuevos flujos de datos, desde constelaciones de sensores atmosféricos hasta la señal de millones de cámaras de seguridad", vaticina Baker. "Es un mundo paralelo que está tomando forma; un laboratorio para la investigación y el descubrimiento compuesto de números, vectores y algoritmos".

## Aplicaciones del 'data mining'

- Comercio. Conocer el patrón de compras del cliente para, por ejemplo, diseñar mejores campañas de publicidad y aumentar las ventas.
- Empleo. Cuantificar las habilidades de los trabajadores para aprovecharlas al máximo. Localizar trabajadores descontentos.
- Elecciones. Cotejar grandes bases de datos sobre hábitos de consumo y encuestas para conocer con detalle las inquietudes y los intereses del votante y así perfilar mejores campañas.
- Medicina. Reunir grandes bases de datos biológicos para emplearlas en la detección a tiempo de tumores, demencias o alzhéimer.
- Lucha contra el terrorismo. Localización de relaciones, transacciones económicas y movimientos de sospechosos.

