

ANÁLISIS DEL BIG DATA

EL VALOR DE L

El Big Data está presente en todos los sectores: transporte, sector de seguridad, seguros, asistencia médica, etc. Sin embargo, los grandes datos no son nada sin un análisis que les de sentido.

La Comisión Europea se ha comprometido a desarrollar la economía del Big Data. La Dirección General de Comunicaciones, Redes, Contenidos y Tecnología, 'DG CONNECT', considera que para incrementar la competitividad europea resulta vital **centrarse en la gestión de la información inteligente y del Big Data.**

"Aquí hay una oportunidad gigantesca", dijo Neelie Kroes, vicepresidenta de la Comisión Europea y responsable de la Agenda Digital Europea, en su intervención en la Conferencia Big Data 2013 realizada en Bruselas. "Los datos pueden revolucionar la manera en que tomamos las decisiones, y pueden resol-

ver muchos problemas, optimizar la producción, la gestión de la asistencia sanitaria o la de los recursos en general".

El torrente de información que se produce en torno a la actividad de cualquier ser humano y de cualquier máquina del planeta, desde los tweets a las cáma-

no es fácil utilizar los métodos de análisis convencionales.

En según lugar, se trata de grandes cantidades. Nadie sabe exactamente cuántos son, pero se ha estimado que el 90% de los datos existentes ahora mismo se han generado durante los dos últimos años.

tor científico de Thales Communications & Security. "En la actualidad, la manera de almacenar los datos está cambiando y tenemos que reinventar la administración de todo el proceso para gestionar los datos directamente, con una nueva manera de realizar los cálculos. Por

Con la gran cantidad de datos disponibles que hay en la actualidad, quizás no sea tan sorprendente que la fiabilidad de estos aparezca como nueva frontera

EN POCAS PALABRAS

1 El Big Data como idea se ha grabado en la imaginación de todos. Sin embargo, son las analíticas de estos datos lo que ofrece posibilidades reales.

2 El análisis de las inmensas cantidades de datos que actualmente están en circulación servirá para transformar todos los sectores, aunque todavía no se conoce el detalle.

3 Es necesario disponer de las herramientas adecuadas para separar el ruido del significado y para determinar las conexiones subyacentes y las relaciones correspondientes, que pueden ser de un valor incalculable.

ras de seguridad, los satélites y los sensores de tráfico, supone un potencial increíble. Esta cantidad gigantesca de datos no se puede extrapolar a partir de pequeñas muestras sin unos conocimientos previos de la población y, en este contexto no es necesario realizar una manipulación exhaustiva de todo el conjunto de datos. Las presuposiciones que ya se han realizado con los datos también pueden ser más fiables: en lugar de presionar los datos para que encajen en las hipótesis (modelo impulsado por la hipótesis), **los datos mismos pueden indicar la realidad** (modelo impulsado por los datos).

El análisis de Big Data

Pero el tratamiento de Big Data presenta grandes retos. En primer lugar, una gran parte de los almacenes de información se refiere a datos estructurales y es cierto que los datos recopilados y almacenados están cada vez más desestructurados, así que

La velocidad también plantea problemas, ya que una cantidad significativa de los datos se genera en tiempo real y, en muchos casos, **hay demasiados para almacenar.** Por esta razón, se presenta también la cuestión de su veracidad: ¿cómo evaluar la calidad de cantidades inmensas de datos no estructurados? ¿Son de fiar?

En respuesta al cambio continuo en este paisaje de Big Data, los análisis están desarrollando un arsenal de nuevas herramientas y técnicas para interactuar con conjuntos de datos grandes y complejos, para generar resultados e ideas.

"En la era de la minería de datos, gestionábamos grandes bases de datos con muchas filas y columnas", nos indica Jean-François Marcotrichino, vicepresidente y direc-

ejemplo, se debe establecer la prioridad en la paralelización y la linealización de los códigos de programación ya existentes y los de nueva creación para la minería de datos, además de la nueva arquitectura de almacenamiento de datos como NoSQL, desarrollada por los grandes protagonistas tecnológicos



Para poder extrapolar esta cantidad gigantesca de datos, son necesarios conocimientos previos de la población. ➔

OS DATOS



de Internet: Google, Amazon, Facebook, etc.”.

Estas técnicas avanzadas de manipulación y analíticas de datos abren nuevos caminos. Uno de ellos es el análisis de las redes sociales, donde se usan **técnicas avanzadas de procesamiento de grandes cantidades de datos** para extraer nuevas ideas de redes sociales como Twitter y Facebook. Los datos ‘gráficos’ que generan estas redes sociales son muy valiosos ya que revelan correlaciones entre tiempo, lugar, identidad y contenido, por ejemplo, palabras clave de los textos, a una escala gigantesca.

Pero no es nada fácil pescar

regularidades en un océano de datos. Por ejemplo, Facebook tiene aproximadamente 2.200 millones de usuarios activos, mientras que Twitter tiene unos 350 millones. A efectos de analíticas, cada usuario se considera un ‘nodo’ y el número de enlaces entre nodos o usuarios, se cuenta por decenas de miles de millones. Reducir toda esta complejidad hasta un nivel manejable es el primero paso.

En primer lugar, identificamos las comunidades. Se están desarrollando herramientas que nos permiten realizar esta operación detectando los vínculos entre personas, sin anali-



Considerando la gran cantidad de información que se genera con el Big Data, la capacidad para analizar su veracidad es cada vez mucho más importante.

zar comentarios ni contenidos específicos, nos explica Marco torchino. “El paso siguiente es seleccionar los nodos que están intercambiando palabras clave para facilitar la tarea de localizar las comunidades que estamos buscando. A continuación, examinamos la semántica de los intercambios para conseguir una mejor imagen de lo que están haciendo realmente tales comunidades”.

Acoplar el Big Data con las analíticas de estos datos **nos permite realizar predicciones y detectar patrones** que de otra manera no se observarían. El respeto de la ley es uno de los campos donde estas ideas y su-

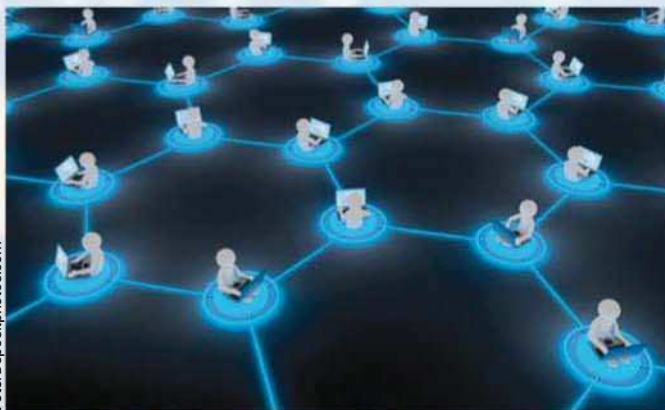
gerencias son de un interés particular, ya que nos pueden ayudar a identificar cualquier forma de terrorismo, alteración social o fraude. Por esta razón, se están desarrollando herramientas que ya son objeto de evaluación por las agencias policiales.

La veracidad de los datos

Además de aumentar la seguridad pública, las estrategias para tratar las analíticas del Big Data pronto podrían ser de utilidad para las empresas, en su tarea de combatir las amenazas a su reputación. Las malas noticias, que a veces se han producido de una manera maliciosa, viajan rápido por las redes sociales. Es necesario disponer de herramientas de limitación de los daños para facilitar que las empresas puedan monitorizar los sentimientos de sus consumidores y responder con rapidez cuando las cosas van mal.

“Los rumores que se propagan por las redes sociales pueden ser muy dañinos para las empresas, principalmente porque el contagio se basa en la inmediatez y la velocidad de la transferencia de datos”, nos dice Marcotorchino. “Tenemos en mente aplicaciones específicas para las empresas, destinadas a analizar los rumores”.

Con tantos datos disponibles como hay ahora, quizás no sea tan sorprendente que la fiabilidad de los datos aparezca como nueva frontera. Considerando

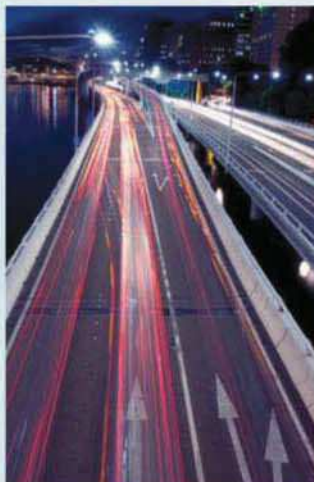


En las redes sociales, cada usuario se considera un ‘nodo’ y el número de enlaces entre nodos o usuarios, se cuenta por decenas de miles de millones.

ANALÍTICAS DE LA CIUDAD INTELIGENTE

La movilidad urbana es una de las áreas donde las analíticas del Big Data podrían pronto marcar una gran diferencia. Las autoridades del transporte público, así como los operadores acumulan enormes cantidades de datos, muchos de los cuales vienen generados por las tarjetas inteligentes de transporte. Por ejemplo, los sistemas de procesamiento de tickets generados por la compañía Thales están instalados en 100 ciudades de todo el mundo y procesan 50 millones de transacciones cada día. Sin embargo, estos datos son únicamente una fracción mínima de su verdadero potencial. Los enfoques que se basan en las analíticas de los Big Data

nos indica la clave para obtener más de los datos existentes y abrir las puertas a un futuro más interconectado, con unas concepciones más profundas acerca del comportamiento de los pasajeros, los flujos de pasajeros y la manera en que los viajeros utilizan las infraestructuras y los equipos. Este tipo de información puede utilizarse para planificar una infraestructura mejor, para introducir nuevos servicios y para proporcionar datos para el pasajero ajustados a un objetivo, para refinar las tarifas e incluso para dirigir la demanda de los clientes, con un grado de certidumbre muchísimo mayor de lo que ha sido posible hasta el momento.



Uno de los aspectos en el que el análisis de los grandes datos centra su actuación es en la movilidad urbana.

que los datos no estructurados se utilizan cada vez más en las aplicaciones críticas, por ejemplo, en las finanzas, en la detección de los fraudes y en la actuación de la policía y de los jueces, la capacidad para ponderar **la veracidad de los datos** se está volviendo cada vez más importante.

“Ahora tenemos que tratar gigantescas cantidades de información y no siempre resulta posible decidir de inmediato si es verdadera o no lo es”, nos indica Marcotorchino. “La confianza es fundamental: la gente no siempre tiene tiempo de verificarlo todo, a veces necesita respuestas simples tipo ‘sí o no’. Si usted dispone de la posibilidad de realizar verificaciones cruzadas con distintas bases

ya no es necesario revisar toda la base de datos para analizar, quizás, millones de líneas de datos”, nos dice Marcotorchino. “Solo necesita una muestra muy grande de 100.000 clientes de los cuales usted dispone de un perfil completo”.

Los enfoques analíticos también se pueden utilizar para determinar qué datos se van generando y facilitar la eliminación de aquellos datos que sean irrelevantes o superfluos. Las prospecciones comerciales con clientes son un caso típico. En una prospección convencional, cada uno de los participantes responde a las mismas preguntas. Las prospecciones de este tipo consumen mucho tiempo para realizar su construcción y es nece-



Foto: Depositphotos.com

Para sacar el máximo partido a grandes conjuntos de datos, una empresa puede analizar los hábitos de compra como para una campaña de marketing efectiva.

Las técnicas avanzadas de manipulación y analíticas de datos abren nuevos caminos

de datos, podrá imaginarse los posibles beneficios que podría conseguir así, pero también podría utilizar la ‘ley estadística de los grandes números’ mediante crowdsourcing, es decir, las opiniones o los datos que provienen simultáneamente de cantidades gigantescas de personas”.

El gran reto: llegar a la verdad

El arte de sacar el máximo provecho a grandes conjuntos de datos conlleva saber en qué medida, grande o pequeña, tenemos que realizar el análisis para alcanzar un objetivo determinado. Una empresa que ya haya realizado, por ejemplo, la segmentación de sus clientes, quizá conozca lo suficiente de sus **hábitos de compra como para redactar una campaña de marketing efectiva**, sin necesidad de partir de cero. Las muestras muy grandes pueden ser suficientes.

“Si dispone de un conocimiento a priori de la población

sario realizar un gran número de preguntas para conseguir un nivel de profundidad que tenga algún interés. Y entonces interviene el riesgo de fatiga de los participantes, donde los clientes marcarán la opción ‘estoy de acuerdo o no estoy de acuerdo’ decenas de veces, pero sin prestar gran atención o bien no realizarán la encuesta de ningún

modo. Los métodos dinámicos de prospección utilizan técnicas analíticas para alejar a los encuestados de la ambivalencia y extraer lo que realmente piensan. Así, la solución para realizar prospecciones con pasajeros de aerolíneas que ha realizado Thales, trabajando conjuntamente con un partner de la industria, utiliza técnicas lógicas y de selección de respuestas para tomar en cuenta muchas preguntas anteriores.

Las prospecciones se realizan utilizando tabletas y la solución conlleva utilizar herramientas sofisticadas de creación informes donde se realiza la tabulación cruzada, la segmentación de los datos y la detección de tendencias. Esto

permite a los operadores de aerolíneas aumentar sus ingresos y conseguir la lealtad de sus clientes, respondiendo a sus necesidades reales.

Soluciones como esta resaltan la manera en que las innovaciones en el campo de la analítica están mejorando la competitividad y también la eficacia. De este modo, las organizaciones de todo tipo pueden **beneficiarse de la revolución de las analíticas**, desde agencias policiales y judiciales que intentan mejorar las tasas de detección hasta bancos y compañías de seguros que combaten el fraude. Y es un campo en el que la pericia de los europeos está jugando un papel cada vez más importante.

“Tenemos que invertir en analíticas para Big Data, no en los Big Data mismos”, resalta Marcotorchino. “Europa no tiene el mismo nivel de pericia en la lectura de bases de datos que los Estados Unidos. Pero en cuanto se refiere a la analítica de los Big Data, la situación es diferente y gracias a instituciones académicas como París VI en Francia y al Imperial College en el Reino Unido, estamos al mismo nivel que el MIT o Stanford de los Estados Unidos, a veces mejor. Europa tiene el potencial necesario para estar bien posicionado en esta esfera”.



Los analistas están desarrollando nuevas herramientas y técnicas para interactuar con conjuntos de datos grandes y complejos, y generar resultados e ideas.

Foto: Depositphotos.com