

Tabla de Contenidos

Introducción a la filosofía del Big Data 1

Introducción a la filosofía del Big Data

Este artículo fue publicado en el blog [Ciencia Sucia](#).

El escenario actual de Internet abre un nuevo paradigma en computación porque a una agrupación de ordenadores interconectados y en funcionamiento permanente no se le puede ya considerar meramente una máquina de Turing (que calcula y se detiene) sino algo nuevo que engulle permanentemente cantidades ingentes de información, la somete, permanentemente también, a cálculos, y elabora predicciones y análisis usando para ello modelos matemáticos y modelos de racionalidad para la toma de decisiones.

Este nuevo uso de la computación se conoce como fenómeno “Big Data” y ha sido aplicado con éxito notabilísimo a realizar predicciones hasta ahora impensables. Por ejemplo, Google lo usa para predecir la [evolución de epidemias](#) mediante el análisis de los términos de búsqueda que los usuarios introducen en su buscador. Quizás el más notable de los éxitos del Big Data haya sido el cálculo político del estadístico Nate Silver que llegó a [clavar los resultados de Obama](#) en las elecciones presidenciales de 2012 usando técnicas de minería de datos propias del Big Data en lugar de las clásicas encuestas sociológicas a las que superó de forma sorprendente en fiabilidad. En su trabajo [La señal y el ruido](#) ha escrito con cierto tono crítico acerca de las posibilidades predictivas del Big Data, indicando que aún no disponemos de buenos métodos por determinar la serie casual de un evento a pesar de haber sobreabundancia de datos.

Sin embargo el siglo XX no parecía augurar este exitoso escenario a comienzos del siglo XXI. A comienzos del siglo XX sucedió una crisis de fundamentación en la matemática como consecuencia de los trabajos de Kurt Gödel que propiciaron el nacimiento de la computación por parte de Alan Turing para demostrar que la incompletitud que Gödel demostró en la matemática tenía como consecuencia la incomputabilidad. Dicho en pocas palabras: no todo es calculable.

Si no todo es calculable ¿cómo es que gracias al cálculo matemático hemos llegado a introducirnos con éxito en el escenario más escurridizo de la matemática, el comportamiento humano?

Pues porque donde decimos “cálculo” en realidad estamos diciendo mucho más. El fenómeno Big Data no hace sencillas sumas y restas con los datos como los que nosotros hacemos al calcular cuánto dinero nos va a quedar a fin de mes. Para hacer una predicción sobre el comportamiento humano los creadores de cálculos Big Data tienen que crear [simulaciones de agentes humanos](#) y ver cómo se comportan en el escenario complejo que analizan. Y aquí es donde está el problema porque no queda claro qué conjunto de valores y prioridades tiene el agente humano simulado en dichos cálculos.

Lo de simular agentes humanos no es nada nuevo. La ciencia económica basa sus predicciones (o más bien [sus intentos por predecir](#)) en un agente humano modelado según la [Teoría de la Acción Racional](#), una simplificación tal de la racionalidad humana que Amartya Sen ha llegado a calificar de imbécil social a la simulación usada por la economía que describe un ser capaz de hacer un cálculo de utilidad sobre qué le conviene, es decir, de aplicar una racionalidad instrumental, pero incapaz de relacionarse con otros de modo cooperativo, algo que sucede permanentemente. El agente humano simulado por la economía es un fracaso. No recoge la diversidad del comportamiento humano.

Pues bien, la misma matemática que Gödel nos enseñó que no era completa, los mismos ordenadores que Turing nos enseñó que no eran capaces de calcularlo todo y la misma teoría racional que Amartya Sen ha calificado como una imbecilidad se han combinado en el fenómeno Big Data para

crear unos modelos de simulación de la conducta social humana con un increíble poder predictivo. ¡Bingo! Con piezas defectuosas hemos construido un Fórmula-1 que gana carreras. Donde la ciencia económica con su deficiente modelo de racionalidad ha fallado parece que el Big Data está acertando: la primera es incapaz de predecir nada y la última parece ser capaz de predecirlo todo.

¿Qué ha cambiado para que ahora sí se pueda calcular la conducta social humana? ¿Es sencillamente asunto de que analizando más datos llegamos a mejores conclusiones o es que se han aplicado mejores modelos de racionalidad a los cálculos Big Data?

Nathan Jurgenson en su reciente artículo [View From Nowhere](#) analiza esas preguntas y acusa al fenómeno Big Data de volver a resucitar al monstruo que creíamos muerto y enterrado de la “objetividad universal”, es decir, la supuesta existencia de una metodología neutra que sin ningún tipo de sesgo analice fríamente petabytes de datos y produzca certeros análisis con la misma naturalidad con la que el gusano produce la seda. El pensamiento dominante en algunos científicos de datos supone una vuelta al [Positivismo](#), una vuelta a olvidar los sesgos sociales descubiertos en la ciencia por [Thomas Kuhn](#) y también un retroceso al negar la [carga teórica de la observación](#).

Además del peligro de un nuevo positivismo que indica Jurgenson, esta nueva tendencia supone, desde mi punto de vista, dos recaídas en sendos temas que parecían más o menos superados como consecuencia del pensamiento desarrollado en el siglo XX y que explicaré en futuras entradas de este blog. Por una parte lo racional vuelve a desplazarse desde el ámbito ontológico al epistémico tal como en su día denunció Heidegger (por eso la informática nace con Turing como ciencia y ahora mayoritariamente se habla de ella como tecnología). Por otra parte una racionalidad surgida en el siglo XX de base débil y que permite recoger todos los matices y la diversidad humana es sustituida en el escenario Big Data por un pensamiento analítico totalizador con indiscutible éxito predictivo pero con nuevas consecuencias alienantes.

Por si éramos pocos, la tendencia que denominamos “Internet en la nube” ha llegado al Big Data y ya se empiezan a ofrecer a la sociedad complejas tecnologías de análisis de datos como productos fácilmente consumibles ([analytics as a service](#)). Toda la familia de los “analytics” (desde [Google](#), a [Facebook](#) pasando por [Twitter](#)) permite interpretar con unos pocos clicks los gigabytes de información recopilada de nuestras interacciones sociales y ofrecérsola como resultados con un aspecto científico impecable e ilustrado con hermosos diagramas de todo tipo.

Y de este pastel los smartphones son la guinda. Dotados con cada vez más cantidad de sensores nos empiezan a permitir recopilar datos de nuestro entorno, incluso de nuestras constantes vitales, por lo que ya están apareciendo aplicaciones (Apps) que simulan ser un experto capaz de interpretarnos qué significa toda esa información y advertirnos de un riesgo inminente.

Esta generalización del consumo de Big Data solo empeora la situación pues propicia la recepción acrítica por parte del público de esta nueva versión de la vieja racionalidad instrumental. El público ha sido adiestrado para delegar las decisiones en el experto (sea este humano o simulado) y se le convence de la fiabilidad del experto con incontestables diagramas y cifras que muestran una realidad que no tiene capacidad de criticar y acepta dócilmente, convencido de que las “nuevas tecnologías” son logros de cálculo gracias a la potencia de las nuevas máquinas.

Los árboles de la tecnología le impiden al consumidor de estos productos ver la maraña conceptual que se le está presentando aunque sea con una estética impecable. Si el smartphone me dice que se acerca la gripe, que Obama ganará las elecciones o que podría tener un infarto pues eso significa que esa es la realidad, ese es el mundo en el que vivo. Punto y final. Y lo más interesante es que luego la gripe llega, Obama gana y aunque no te dé el infarto se te ha avisado de un riesgo, con lo que la tecnología ha cumplido su misión.

Riesgos, expertos, análisis y predicciones. Un cocktail que puede ser muy venenoso pero con el que tenemos que acostumbrarnos a convivir. Vivimos en la “sociedad del riesgo” como muy bien ha explicado [Ulrich Beck](#), en una sociedad de simulación del conocimiento experto, como ha explicado [Axelrod](#) y en una sociedad generadora de incertidumbre, como explica [Sven Ove Hanson](#). Ante este escenario tan retorcidamente complejo no podemos aplicar un modelo de racionalidad tan simple como el actor racional del modelo económico que tiene capacidad de calcular qué le conviene. Nuestra sociedad es tan compleja que queda fuera de nuestra capacidad calcular qué nos conviene. De hecho probablemente nunca lo hicimos. Es posible que, como comentan [Gerd Gigerenzer](#) y [Jose Francisco Álvarez](#) nuestra racionalidad sea mucho más heurística de lo que nos han contado y completamente acotada: decidimos con información parcial, incompleta y en constante cambio. Decidimos con mecanismos biológicamente creados, no matemáticamente formulados. Hacer un modelo matemático de esa realidad evolutivamente creada puede ser una tarea monumental aunque hay ya intentos de avanzar en esa línea, como los de [Gregory Chaitin](#) y su idea de modelar la vida como software en evolución.

Si modelar la biología es difícil, hacerlo con la racionalidad humana parece una tarea heroica. Por eso es por lo que el Big Data ha escogido aparcar la creación de nuevos modelos de racionalidad y aplicar los conocidos modelos económicos a ingentes cantidades de datos. Y funciona. O parece que funciona. Pero no conocemos cómo ha sido cargado teóricamente el experto que llevamos simulado en nuestro smartphone o que interpreta nuestras interacciones sociales. Tampoco solemos saber cómo se ha construido el riesgo del que se nos avisa. Se nos pide un acto de fe tecnológico que la mayoría de la gente, cegada por la luz celestial de las “nuevas tecnologías” no tiene inconveniente en conceder. Aceptamos como experto al que se nos presenta como tal. Aceptamos que si ese experto dice que tenemos un riesgo es que efectivamente estamos en peligro. Y si el experto dice que hay un 20% de probabilidades de que nos suceda algo no cuestionamos su método sino que pasamos a decidir si aceptaremos ese nivel de riesgo o no. El experto puede fácilmente crear un escenario de consentimiento inducido máxime cuando es una aplicación informática que habla un lenguaje de gráficos y estadísticas. Cautivo y desarmado el ejército ciudadano, el experto ha alcanzado sus últimos objetivos sociales, que aún no sabemos si son describir la realidad, predecirla o construirla. Mientras no conozcamos los modelos de racionalidad aplicados en las simulaciones expertas de la conducta humana no sabremos si los supuestos exitos predictivos son logros de la ciencia social o [profecías autocumplidas](#).

Esta situación que estamos describiendo queda perfectamente explicada en el número de junio de 2008 de la revista Wired titulado The End of Science (El final de la Ciencia) donde en el artículo [The Petabyte Age](#) se indica que estamos ante un nuevo paradigma que no es de base experimental ni idealista sino una especie de tercera vía de naturaleza computacional:

“Scientists have always relied on hypothesis and experimentation. Now in the era of massive data, there’s a better way”, o “Solving scientific problems used to require grand theories. Now it’s just a matter of crunching the numbers.”

Los Estados no han sido ajenos a los éxitos predictivos de estas tecnologías y a sus usos como mecanismos de intervención y control social estableciendo [programas secretos de recopilación ilegal de datos personales](#) a sabiendas de las posibilidades que ofrece su análisis. Esto abre el camino a nuevos estudios sobre las relaciones de poder en este escenario de nueva racionalidad computacional de carácter positivista.

Por otra parte, ¿podemos seguirnos permitiendo el lujo de educar a la gente solo en competencias tecnológicas pensando que la mayoría de ellos no quieren pensar y lo que quieren es solamente una App que les solucione los problemas?

Si convertimos al ciudadano en mero consumidor de tecnología, incapaz de criticar los métodos que se usan para hacer analítica de datos estaremos repitiendo el escándalo sucedido en la economía con la fórmula [Black-Scholes](#), una fórmula que parecía capaz de predecir el comportamiento de los “derivados”, esos nuevos tipos de productos financieros que en principio produjeron un crecimiento exponencial de la actividad económica y que, como era una descripción falsa de la realidad, produjo el colapso financiero que vivimos desde 2008. La fórmula matemática Black-Scholes fue la justificación matemática de una teoría que se probó falsa. Modelar matemáticamente el comportamiento de un producto financiero o de una sociedad no conduce necesariamente a una descripción certera de su comportamiento.

Considero necesario solucionar la falta generalizada de crítica sobre la racionalidad aplicada a la interpretación de los datos masivos elaborada por la nueva ciencia de datos a la que denominamos Big Data, así como iniciar programas específicos de investigación sobre dicha racionalidad de modo que se pueda contrarrestar un nuevo positivismo de corte computacional.

Para dicha crítica ya está abierto un fundamento filosófico que exploraré en futuras entradas de este blog.

BIBLIOGRAFÍA

[Crisis de la razón y racionalismos contemporáneos](#)

[Decisiones tecnológicas: racionalidad acotada, racionalidad ecológica, heurística](#)

[Lo social y lo racional](#)

[Gerd Gigerenzer](#)

[Teorías de la racionalidad - José Francisco Álvarez](#)

[El cambio técnico: racionalidad, información, riesgo e incertidumbre](#)

[Decisiones tecnológicas: sistemas expertos, agentes artificiales y simulación](#)

From:

<http://filosofias.es/wiki/> - **filosofias.es**

Permanent link:

<http://filosofias.es/wiki/doku.php/cts/bigdata/introduccion>

Last update: **2016/11/22 11:04**

