14 de agosto de 2019

El economista Walter Sosa Escudero, sobre la profusión de datos masivos e instantáneos

La cocina del Big Data

¿Se puede predecir una elección presidencial o el podio de un mundial de fútbol? Este investigador matiza el énfasis revolucionario del fenómeno y afirma que la verdadera transformación vendrá cuando las personas generen nuevas preguntas.

Por Pablo Esteban

Tenemos, como nunca, un montón de información disponible y ello nos entusiasma pero también nos obnubila. Por un lado, el futuro se aproxima como un auténtico paraíso; lo que antes valía horas de bibliotecas, hoy se resuelve en cuestión de segundos. Sin embargo, tanta potencia tecnológica tiene su reverso: accedemos a muchos datos pero a cambio dejamos los nuestros. Cuando colocamos una dirección en el GPS, enviamos un correo o cenamos con amigos en algún bar, imprimimos nuestras huellas en el ciberespacio. Nos tornamos geolocalizables aunque nadie nos busque y nos volvemos predecibles aunque poca gente aprecie nuestra intimidad. La era global es así, nos bambolea entre el confort y el pánico, en cualquier tiempo y sin importar el lugar. En los intersticios, se teje una gran transformación cultural que sabemos que existe pero todavía no somos capaces de poner en palabras. Lo bueno es que no estamos solos, tenemos jinetes que custodian el terreno, utilizan sus largavistas y nos ayudan a comprender las cosas en su justa medida y por anticipado. Walter Sosa Escudero es economista (UBA), profesor de la Universidad de San Andrés e Investigador del Conicet. Acaba de publicar “Big Data” por la editorial Siglo XXI y coloca paños fríos a una situación que aparenta volar de fiebre.

--¿Qué es el Big Data?

--Lo primero, lo rápido, lo Wikipedia, sería definirlo como un fenómeno de profusión de datos masivos e instantáneos; producidos de manera espontánea y de expresión anárquica a través de sistemas que están interconectados entre sí. En la actualidad, a menudo, pensamos que cuánto más datos relacionamos es mejor y no siempre es así. No se pueden comparar los datos que arrojan 1000 encuestas bien equilibradas, o bien, los 60 que se manipulan en un experimento

de laboratorio elaborado de manera concienzuda, con los 300 mil que surgen de responder a un multiplechoice en twitter. Para no celebrar el Big Data como si de una revolución se tratara, me gusta decir que lo importante no es trabajar con muchos datos sino que estén bien curados. Seguramente, serán mucho más informativos que aquellos extraídos sin ningún criterio.

--¿Cómo utilizar tanta información a favor?

--Hablar de Big Data se parece mucho a hablar de cerveza artesanal. El reverso de los datos, históricamente, fueron los métodos que se emplean para estudiarlos. Cuando al triángulo de datos-métodos-ideas se le quita una pata, éste se desploma. Tener más información es una buena noticia siempre y cuando existan preguntas que reclamen la necesidad de ser respondidas. En el futuro, la condición de posibilidad para una revolución de los datos tendrá que ver, sobre todo, con la capacidad de las personas para generar preguntas y discutir ideas.

--¿No hay preguntas que reclamen ser respondidas?

--Hay pero no en la proporción en que hubiésemos querido. En los últimos 15 años hubo una explosión de datos y métodos, pero no de ideas. Predecir quién ganará un mundial o quién triunfará en las próximas elecciones sigue siendo igual de difícil que siempre. En las épocas de los algoritmos y del Big Data, no estamos realizando mejores predicciones deportivas ni políticas. Basta con ver lo que ocurrió con la emergencia de Donald Trump, el derrotero del Brexit y, sin ir tan lejos, la aparición de la fórmula Fernández-Fernández. Con machine learning de por medio, nadie vio la jugada. Nadie. Lo que no advertimos es que fenómenos tan complejos como la economía o la política poseen una dificultad intrínseca que no tiene nada que ver con la falta o el exceso de datos.

--Si bien los datos no pueden predecir grandes fenómenos, son útiles para inducir decisiones. De hecho, el mercado publicitario se estructura de ese modo.

--Absolutamente. Si me preguntaras cuál es el mayor rédito de todo esto, te diría que reside en la capacidad inductiva, descriptiva, ordenadora y clasificatoria del Big Data. En el pasado uno hablaba de “el votante” y hoy se puede trabajar con una granularidad muchísimo más grande: “el votante que tiene entre tantos y tantos años, que vive en el barrio x con su pareja y asiste a la universidad x”. En la publicidad ocurre lo mismo, hubo algo por lo que McDonald’s se diversificó y existen McCafés; evidentemente, alguien con buen ojo se dio cuenta de que los chicos eran llevados por padres que preferían tomarse un café después del cine y no comer un combo.

--Ahora que sabemos sus potencialidades, ¿para qué no sirve el Big Data?

--No sirve para explicar, porque eso es mucho más difícil. Mucha gente piensa que como se puede clasificar mejor, estamos en mejores condiciones de explicar. Un ejemplo burdo: si puedo conseguir los datos de ventas de paraguas en diferentes regiones, podría llegar a estimar cuáles son los puntos geográficos en los que más llueve. Si se venden más en Rosario que en Antofagasta (Chile) debe ser que hay más precipitaciones en la primera que en la segunda. Ahora bien, esto no ayuda a entender por qué llueve ni tampoco contribuye a resolver una sequía o una inundación. No podría asesorar a los que comercian paraguas ni tampoco a los que quieren cultivar la tierra. Como no me permiten entender las razones, entonces no puedo cambiar la realidad. Hay un trecho enorme entre medir la pobreza y entenderla, diseñar una política y evaluar su impacto favorable. Los datos podrían funcionar como una herramienta para ganar una elección, pero no son definitivos porque requieren de un trabajo interpretativo y creativo de los analistas.

--Con la exposición de nuestros datos en redes sociales vivimos cada vez más controlados. ¿Perdimos libertad, existen nuevas formas de libertad, o bien, la libertad nunca existió?

--A los efectos de no sembrar rastros, cada vez que salgo al trabajo debería dejar el celular en mi casa. Pero eso me complica la vida; no puedo chequear el tránsito, no sé si me conviene moverme en medios de transporte públicos, usar

la bici o ir caminando. Tampoco tengo datos del tiempo atmosférico y no puedo saber cómo vestirme. Al acceder a los servicios que brinda internet, me beneficio de un montón de cosas y alguien cobra por ello. Como pasa con cualquier tecnología, el asunto es hacer la cuenta entre las ventajas y las desventajas. El obstáculo con el Big Data y la capacidad predictiva de los algoritmos es que hoy ese balance no arroja un resultado certero. El límite con la privacidad individual se sostiene en un terreno fangoso. No me gustaría enterarme por un robot cosas del tipo: “hoy estuviste tomando un café con un amigo y muy probablemente tiene cáncer”. El sistema puede saber mucho de nosotros, incluso, podría saber cosas que ni siquiera sabemos nosotros mismos de nosotros.

--El asunto requiere de un abordaje matemático y estadístico, pero también posee ribetes filosóficos...

--Sí, porque nos lleva a pensar cuestiones que creíamos olvidadas. El Big Data está a mitad de camino entre la revolución y convertirse en más de lo mismo. Con las redes sociales sucede parecido: nos plantean un nuevo modo de interacción, pero ello no quiere decir que antes no nos relacionábamos. La chica del barrio que te guiñaba un ojo cuando salía a andar en bici hoy es el “Me gusta”. Siempre hubo mecanismos de validación social; prefiero creer que los vínculos se tejen de otra manera. Hoy se habla de las burbujas informativas como una novedad, es decir, le pongo *likes* a los que piensan como yo y borro a los que lo hacen distinto. El diario de Irigoyen era un poco eso. La moraleja: seamos cautos con comunicar revoluciones, especialmente, cuando todavía no suceden.

poesteban@gmail.com