lograr sorprendentes análisis respecto de las percepciones, sentimientos, opiniones que tienen los internautas sobre un producto o servicio en tiempo real. El problema es que solo en Twitter se están generando más de 40.000 tweets por segundo, cifra que irá en aumento en los próximos años [1].

REFLEXIÓN FINAL

Hay algo muy interesante en casi todos los problemas relacionados a Big Data: los algoritmos y técnicas que se desarrollan no son para un tipo de dato en específico. Dicho de otra manera, si desarrollamos un algoritmo de análisis de serie de tiempo para datos generados en un radiotelescopio, y luego lo aplicamos a los generados por un electroencefalograma, con algunos ajustes claro está, podremos extraer patrones a partir de las ondas cerebrales de un paciente y quién sabe, detectar en forma temprana una anomalía. Como en todo nuevo concepto, hay mucho de mito y poco de realidad. Aparecen miles de expertos, gurúes del área, pero que están igual que todos nosotros: somos testigos del nacimiento de algo grande que comienza recién a dar sus primeros pasos en ciencia, en tecnología y en los negocios.

Big data: ¿la misma cerveza pero con otro envase? Aún no lo tenemos claro, pero conviene que esta vez se la tome con calma y muy helada, sino puede causarle una big indigestion. ■

REFERENCIAS

[1] "Detecting Trends on the Web: A Multidisciplinary Approach", Dueñas Fernández R., Velásquez, Juan D. and L'Huillier, Gastón. Information Fusion, 20:129-135, 2014.

[2] "Web Mining and privacy concerns: Some important legal issues to be consider before applying any data and information extraction technique in Web-based environments", Velásquez, Juan D., Expert Systems with Applications, 40 (1): 5228-5239, 2013.

[3] "A neurology-inspired model of web usage", Pablo E. Román and Juan D. Velásquez, Neurocomputing, 131: 300-311 ,2014.

[4] "Are you ready for the era of 'big data'", B. Brown, M. Chui J. Manyika, McKinsey Quarterly, 4:24-35, 2011.

[5] "Big data: the management revolution", A. McAfee and E. Brynjolfsson, Harvard Business Review, 90(10):60-68, 2012.

LA NUEVA ERA DE DATOS EN ASTRONOMÍA

FAVIOLA MOLINA

Investigadora postdoctoral en el Departamento de Ciencias de la Computación, Universidad de Chile, donde trabaja con el profesor Alexandre Bergel. Doctora en Ciencia de la Universität Heidelberg y fellow del Instituto Max Planck para Astronomía en Heidelberg, Alemania (2013). Astrónoma de soporte en el Observatorio Europeo Austral (2008-2009). Magíster en Astronomía y Astrofísica de la Pontificia Universidad Católica de Chile (2008). Licenciada en Física de la Universidad de Los Andes, Venezuela (2004). Áreas de interés: Astro-computación, análisis estadístico del medio interestelar y formación de estrellas, poblaciones estelares y recientemente formación de discos planetarios y de transición.

fmolina@dcc.uchile.cl

El desarrollo de cualquier disciplina científica involucra el manejo de datos. En el caso particular de la Astronomía, el incremento de la cantidad y tamaño de los archivos de datos ha ido creciendo con el paso de los años, considerando así a ésta como una ciencia de datos intensivos [Hassan and Fluke, 2011].

Hasta mediados del siglo XX, en específico en la Astronomía óptica, los detectores eran placas fotográficas. La exposición de las mismas podía tomar horas de acuerdo a la intensidad del objeto que se quería observar. Más tarde, se dio paso a los fotómetros fotoeléctricos que ofrecían mayor sensibilidad, precisión, linealidad y un mayor rango dinámico para el análisis que las placas fotográficas[1].

1 http://star-www.rl.ac.uk/docs/sc5.htx/node7.html

El curso de los datos astronómicos cambió dramáticamente cuando en 1975, dadas las mejoras tecnológicas en los métodos de recolección de imágenes, se comenzó a proponer la idea de implementar los dispositivos de carga acoplada (CCD por su siglas en inglés[2], Charge Coupled Device) en la obtención de datos [Samuelsson, 1975; McCord and Bosel, 1975]. Como resultado de estas propuestas, en 1976, los CCDs revolucionaron la astronomía cuando J. Janesick y B. Smith adosaron un CCD a un telescopio de 155 cm. de diámetro (localizado en el Monte Bigelow, Arizona) para obtener imágenes de Júpiter, Saturno y Urano (Parimucha and Vanko, 2005). La ganancia en sensibilidad fotométrica y cobertura espectral a partir de esa época ha ido aumentando drásticamente.

Hoy en día, la cantidad de datos astronómicos (tanto observacionales como teóricos) en archivos sobrepasan los Petabytes[3] [Hassan and Fluke, 2011]. El término "big-data" se ha implementado para describir sets de datos que son muy grandes para ser manejados con las herramientas de procesamiento, análisis, y visualización, existentes a la fecha. Con el paso del tiempo, la cantidad de observatorios obteniendo datos experimentales ha crecido, así como los centros de investigación teórica que producen grandes cantidades de datos simulados. Con la nueva generación de telescopios e instrumentos, la resolución espacial y espectral de las imágenes ha aumentado de una manera sin precedentes. Junto con esto, el incremento en la capacidad de cómputo y almacenamiento ha provocado que el tamaño de cada archivo de datos crezca significativamente. Por ejemplo, cuando el Atacama Large Millimeter/sub-millimeter Array (ALMA) se encuentre en completa operación, generará más de 750 Gb de datos por día[4], que se traduce en unos 250 Tb por año[5]. Otro ejemplo son las simulaciones cosmológicas con las cuales se resuelven numéricamente las ecuaciones que rigen la dinámica del Universo. Estos códigos usan del orden de 10^{10} partículas y/o grandes mallas adaptativas tridimensionales en el computo que producen archivos de datos de varios Terabytes, incluso pudiendo llegar al orden de los Petabytes [6] [Springel et al., 2005].

Por otro lado, los datos astronómicos cada vez están más interconectados. Dada la cantidad de observaciones realizadas durante las últimas décadas, y la recopilación y digitalización de datos históricos (en distintas bandas del espectro electromagnético), ha nacido el Observatorio Virtual (The Virtual Observatory, VO: http://www.ivoa.net/). De este gran repositorio es posible obtener datos que posibilitan estudios sistemáticos, pancromáticos y estadísticamente significativos de la evolución de sistemas astronómicos [Brunner et al., 2002]. El VO alberga datos provenientes de archivos de una gran cantidad de observatorios terrestres y espaciales, entre ellos están el Telescopio Espacial Hubble (HST), el Observatorio de Rayos X Chandra, El Sondeo Digital Sloan del Cielo (SDSS), el Sondeo de Todo el Cielo en Dos Micrones (2MASS), el Sondeo del Cielo Digitalizado del Observatorio Palomar (DPOSS), el Observatorio Europeo Austral (ESO), el Telescopio para el Sondeo de Astronomía Visible e Infrarroja (VISTA), y más recientemente ALMA, además de muchos otros. Esta nueva manera de hacer astronomía permite el uso de cualquier tipo de datos a científicos y estudiantes desde cualquier parte del mundo, y que antiguamente no tenían acceso a grandes observatorios.

La cantidad de datos promete continuar incrementándose estrepitosamente con el paso del tiempo. En el futuro cercano, nuevos telescopios tales como el Telescopio Sinóptico Grande para Sondeos (LSST)[7], el Telescopio Magallanes Gigante (GMT)[8], el Telescopio de Treinta Metros (TMT)[9], el Square Kilometer Array (SKA)[10], el Telescopio Europeo Extremadamente Grande (E-ELT), entre otros; abrirán nuevas posibilidades en la Minería de Datos.

En esta nueva era astronómica es importante notar que, no solo es necesario contar con grandes infraestructuras de almacenamiento y altas velocidades de transferencia, sino también que es prioritario desarrollar herramientas de visualización que posibiliten realizar análisis de manera ágil y rápida. Actualmente, en conjunto con el Profesor del Departamento de Ciencias de la Computación (DCC) de la Universidad de Chile, Alexandre Bergel, nos encontramos desarrollando una plataforma dinámica y diligente para la visualización y análisis de datos astronómicos: AstroCloud (http://astrocloudy.wordpress.com). Esta herramienta está enfocada en el análisis específico de un problema físico determinado, pero con la flexibilidad de agregar fácilmente nuevos módulos que incluyan distintos tipos de análisis. Por otro lado, la nueva estructura de datos astronómicos requiere de una adecuada visualización y análisis tridimensional[11]. AstroCloud será una herramienta que facilitará y reducirá el tiempo empleado en el análisis de datos masivos en dos y tres dimensiones. Ésta es justamente una de las necesidades principales de observatorios como ALMA. Actualmente contamos con la colaboración de la Profesora Nancy Hitschfeld (DCC U. de Chile), el Profesor Lucas Cieza (Núcleo de Astronomía, Facultad de Ingeniería, Universidad Diego Portales), la Dra. Paola Pinilla (Observatorio de Leiden, Holanda), así como con la ayuda de Juan Cortés (ALMA) quien ayudará a probar la versión beta de la herramienta. Este proyecto multidisciplinario está abierto a estudiantes entusiastas que deseen participar en el desarrollo de esta dinámica herramienta. ■

2 En lo sucesivo, todas las siglas y acrónimos refieren a las siglas en inglés correspondientes a la expresión.
3 1015 bytes.
4 El tamaño de cada imagen depende del modo de observación.
5http://www.almaobservatory.org/en/press-room/announcements-events/542-virtual-observatories-chilean-developmentof-astronomical-computing-for-alma
6 e.g., http://www.cfa.harvard.edu/itc/research/movingmeshcosmology/
7 http://www.lsst.org/lsst/
8 http://www.gmto.org
9 http://www.tmt.org
10 http://www.ska.ac.za/about/project.php
11 En Astronomía, las tres dimensiones son dos en posición-posición que corresponde al plano del cielo, y la tercera (profundidad) corresponde al espacio de frecuencias y/o velocidades.

REFERENCIAS

Brunner, R. J., Djorgovski, S. G., Prince, T. A., and Szalay, A. S.: 2002, in J. S. Mulchaey and J. T. Stocke (eds.), Extragalactic Gas at Low Redshift, Vol. 254 of Astronomical Society of the Pacific Conference Series, p. 383.

Hassan, A. and Fluke, C. J.: 2011, Publications of the Astronomical Society of Australia. 28, 150.

McCord, T. B. and Bosel, J. P.: 1975, in Charge-Coupled Device Technology for Scientific Imaging Applications, pp 65-69.

Parimucha, S. and Vanko, M.: 2005, Contributions of the Astronomical Observatory Skalnate Pleso 35, 35.

Samuelsson, H.: 1975, ESA Scientific Technical Review 1, 219.

Springel, V., White, S. D. M., Jenkins, A., Frenk, C. S., Yoshida, N., Gao, L., Navarro, J., Thacker, R., Croton, D., Helly, J., Peacock, J. A., Cole, S., Thomas, P., Couchman, H., Evrard, A., Colberg, J., and Pearce, F.: 2005, Nature 435, 629.

MANEJO DE DATOS MASIVOS EN BIOMEDICINA COMPUTACIONAL

VÍCTOR CASTAÑEDA

Ingeniero eléctrico, Universidad de Chile (2005) y Doctor en Ciencias de la Computación, TUM-Alemania (2012), especializado en el procesamiento de imágenes médicas. Su Tesis Doctoral estuvo enfocada en el procesamiento de imágenes endoscópicas y sensores 3D tales como cámaras Time-Of-Flight y Kinect. Actualmente hace su postdoctorado (Proyecto FONDECYT 3140444) en el seguimiento de núcleos y segmentación de membranas celulares provenientes de microscopía Light Sheet, en SCIAN Lab del Prof. Härtel, ICBM, Facultad de Medicina, U. de Chile.

vcastaneda@med.uchile.cl

A nivel mundial, la investigación de excelencia en el ámbito biológico, clínico, médico y biomédico depende en forma crucial de la capacidad de análisis de los datos recolectados por experimentos que generan crecientes volúmenes de datos. Como ejemplo, la microscopía confocal *in vivo* es capaz de generar cientos de gigabytes (GB) de imágenes tridimensionales en una sola captura. En estos casos los investigadores de BioMedicina Computacional recurren a información cuantitativa, mediante modelamiento matemático y computacional para entender y predecir procesos biológicos con relevancia en medicina y ciencia básica.

Las aplicaciones involucradas entrelazan a los mundos de la Salud Pública (bases de datos), Clínico/Hospitalario (sistemas de información), (Neuro) Ciencias Biomédicas (imágenes, bioinformática y biología computacional), Ciencias de la Computación/Ingeniería (algoritmos), Física y Matemática (herramientas y modelos). Se reconoce que la creación del campo de la BioMedicina Computacional requiere un esfuerzo mayor a través de los años para generar equipos multidisciplinarios y una nueva cultura de trabajo desde las ciencias básicas hasta la investigación clínica, salud pública, y la introducción de nuevos servicios en sistemas de salud. Hasta la fecha, Chile y la mayoría de los países latinoamericanos no cuentan con respuestas adecuadas en este tema, principalmente por la falta inversión de fondos estratégicos con visión de mediano y largo plazo. La obtención de esta capacidad conlleva a la creación de alianzas estratégicas entre las disciplinas involucradas para desarrollar y acceder a nuevas tecnologías necesarias para el análisis de datos masivos. Instituciones emblemáticas en todo el mundo han respondido a este desafío a través de la creación de centros o institutos que persiguen misiones afines: (i) Johns Hopkins University [1], (ii) U-Michigan [2], (iii) U-Cincinnati [3], (iv) Janelia Farm [HHMI] [4], (v) BioQuant, DKFZ, Uni-Heidelberg [5], y (vi) Mt. Sinai [6], por nombrar solo algunos.

El análisis de datos en BioMedicina Computacional requiere de una

→