

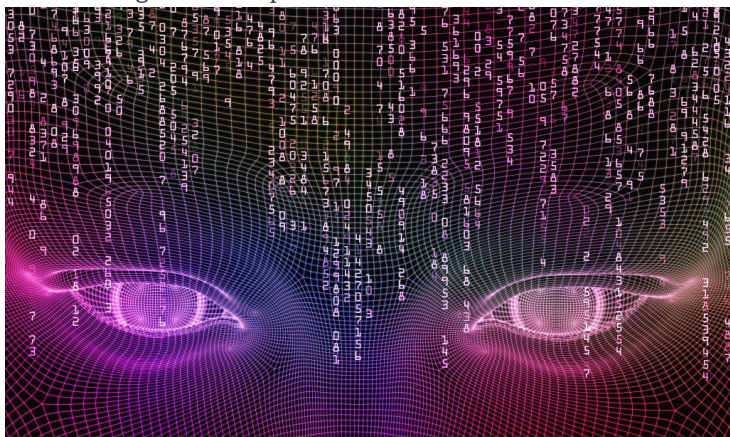
Big Data en medicina: La inteligencia artificial que viene.

Anthony Chang*

Resumen: El artículo ofrece una visión general sobre lo que es y significa el Big Data así como su aterrizaje en la biomedicina: sus posibles riesgos, beneficios y retos a la hora de su aplicación en la atención sanitaria.

Big data in medicine: The upcoming artificial intelligence

Abstract: The article offers an overview of what Big Data is and what it means its landing in biomedicine: possible risks, benefits and challenges when implement in health care.



Big Data es el más reciente paradigma de datos que significa no solo una información que es mucho más voluminosa que cualquier otra forma de almacenaje estructurado de datos, sino también el cruce de esta masiva cantidad de información con análisis de datos sofisticados al objeto de adquirir nuevos conocimientos o visiones [1]. Este fenómeno ha tenido lugar en sectores tales como los negocios [2], las finanzas [3], o los deportes [4]. En las últimas elecciones presidenciales en los EEUU, el científico de datos Nate Silver demostró que su metodología analítica de datos era superior a las opiniones de muchos expertos políticos para predecir el resultado de la elección [5]. Big Data es, por lo tanto, un movimiento de datos dominante y una revolución del conocimiento que enfatiza que "la información de los datos es más confiable que la intuición de los expertos" y se encontraría ya en el pico descendente "de expectativas infladas" del ciclo de Gartner para las tecnologías emergentes con una estimación 5 años su maduración y adopción (los interesados pueden ver esto en : (

<http://www.mediabuzz.com.sg/asian-emarketing-latest-issue/210-asian-e-marketing/digital-marketing-trends-a-predictionsweek-1/2504-gartner-hype-in-2015-around-the-internet-of-things-iot-and-wearables>)

El big data tienen importantes dimensiones y todas ellas comienzan convenientemente con la letra "v"

(<http://www.datasciencecentral.com/forum/topics/the-3vs-that-define-big-data>):

1) **Volumen** - se calcula que los volúmenes de datos recientes están en petabytes (10¹⁵) y exabytes (10¹⁸) y estarán en zettabytes (10²¹) en un futuro cercano (la unidad más alta es el yottabyte que es 10²⁴) (

<http://www.theopenstrategist.com/2012/10/big-data-growthchart.html>). Estos datos incluyen registros médicos electrónicos, dispositivos de monitorización del hogar, datos genómicos, reclamaciones a seguros, información sobre medicamentos, y datos de imágenes (de ecocardiogramas, angiogramas y RMN/TAC); 2) **Variedad** ? la consideración de los datos estructurados y no estructurados en numerosas formas y combinaciones; y 3) **Velocidad** - la transferencia de datos y el análisis ahora deben estar en fracciones de segundos e incluso en tiempo real, especialmente con transferencia de estudios de imágenes en telemedicina. Unas "v" adicionales incluyen también: 4) **Veracidad**: la precisión y confiabilidad de big data y sus análisis acompañantes y finalmente 5) **Valor** - determinación de cuánto vale esta información para la institución y el usuario.

El Big Data biomédico

El big data biomédico actual, acumulado por registros médicos electrónicos y el archivo de imágenes digitales (alrededor de 20 megabytes o MB por imagen), alcanza la asombrosa cifra de 100-250 exabytes con una tasa de crecimiento anual de 1.2 a 2.4 exabytes [6]. Este big data biomédico, sin embargo, está aún muy fragmentado y desorganizado. Nuestros datos tradicionales con un

enfoque de "arriba hacia abajo" hasta ahora implica o bien 1) unas bases de datos sobre salud o registros (eso implica el ingreso manual de datos con sus limitaciones inherentes de precisión e integridad, seguido de análisis de datos con herramientas estadísticas básicas) o bien 2) una investigación convencional impulsada por hipótesis y ensayos controlados aleatorios (ECA) que se han vuelto costosamente prohibitivos, de un alcance limitado y, a menudo, sin respuestas definitivas a las preguntas clínicas planteadas.

Recientemente, este nuevo paradigma de big data se ha aplicado con éxito a la ciencia biomédica principalmente en forma de medicina genómica y su escalada de big data en la transcripción genética [7]. La gran magnitud y la adquisición rápida de este big data genético es increíblemente vertiginosa, como ejemplifica Michael Snyder, un genetista de Stanford que ha generado 30 terabytes de datos con solo sus propios datos biológicos. A pesar de este desalentador desafío, algunos se han enfrentado al mismo con éxito, avanzando en el impacto positivo en la atención del paciente [8] [9]. La piedra angular de todo este esfuerzo de transformación de datos en medicina genómica es el **ENCyclopedia Of DNA Elements (proyecto ENCODE)**, un proyecto internacional de colaboración entre grupos de investigación financiados por el Genome Research Institute con el objetivo de delinear la totalidad de elementos funcionales codificados en el genoma humano [10].

Interpretación del Big Data

El nuevo paradigma de datos en la atención sanitaria será una **información con estrategia de gestión de "abajo hacia arriba"** que implicará un proceso de tres pasos tras la adquisición de los datos: 1) **extracción de datos** con varios almacenes que los proporcionan; 2) **transformación de datos** con datos configurados para un tipo de formato uniforme; y 3) **carga de datos** con los datos ingresados en un sistema de análisis con análisis final. Un ejemplo de esta estrategia de datos fue el seguimiento de la epidemia de gripe que hizo Google (llamado "tendencias de la gripe") y que ha sido perfeccionado en estudios epidemiológicos utilizando grandes conjuntos de datos [11]. El uso de big data con recolección y análisis en tiempo real es facilitado por la nueva tecnología de base de datos Hadoop que minimiza el ciclo de procesamiento de datos de **extracción-transformación-carga** (o ETL) antes mencionado [12].

Big Data puede utilizar estratégicamente metodologías de **inteligencia artificial** para obtener más información esencial e incluso descubrir nuevo conocimiento [13] [14]. El Laboratorio de Informática e Inteligencia Artificial del MIT (CSAIL) lidera algunos de los actuales esfuerzos para combinar Big Data e inteligencia artificial para desarrollar nuevas técnicas en la interpretación de big data que conducirán a una información sinérgica de medicina genómica-clínica. El advenimiento de metodologías de inteligencia artificial como el **procesamiento del lenguaje natural y la computación cognitiva** podrían permitir la administración de datos médicos para no solo organizar sino también extraer datos incluso de registros electrónicos no estructurados automáticamente (sin el tedioso y poco adecuado proceso de entrada manual). Además, la impresionante capacidad del ordenador para interpretar **el sentido de la palabra desambiguación**, o la capacidad de poner palabras en contexto, fue demostrado por la computadora de IBM Watson durante su paso por el programa Jeopardy! y actualmente se usa de forma efectiva para el análisis de datos de cáncer en Instituto del Cáncer del Memorial Sloan-Kettering.

El valor percibido sobre los datos médicos no debería limitarse a los analistas de datos y científicos en los departamentos de TI de los hospitales, sino que se ampliarían a todos los trabajadores de la salud para crear una **cultura centrada en los datos**. Además, big data no tiene solo que ver con la extracción de información importante de las bases de datos médicas, sino que también exige la ejecución de cambios en la práctica e innovaciones transformativas. La transformación de datos definitiva y el conocimiento médico en el futuro implica bases de datos médicos que estarán "vivos" e "individualizados": datos de investigaciones clínicas en curso y análisis con datos estáticos y dinámicos (a través de sensores) de los propios pacientes. Si están incrustados con técnicas de inteligencia artificial, técnicas convencionales (**extracción de datos**) [15] o nuevas (**aprendizaje automático y aprendizaje profundo**) [16], estas bases de datos finalmente habilitarán la "inteligencia" en determinaciones y decisiones para resultados de atención sanitaria.

Implementación del Big Data

Big data se refiere a la calidad, no a la cantidad de los datos. La implementación de Big Data y su análisis podría implicar un pequeño proyecto en un solo sector de un hospital donde los datos puedan ser más fáciles de localizar. Además, el enfoque debería estar en el problema que puede ser resuelto por big data, y no big data o su análisis por sí mismo. Las habilidades analíticas incluso a un nivel básico no deberían poseerlas los analistas de datos, sino deberían tenerlas todos. El proyecto debe ser sometido a mediciones y análisis repetitivos para perfeccionar la estrategia de análisis de datos. Finalmente, la ejecución de un proyecto debe seguir una línea temporal y un progreso uniforme a lo largo del proyecto. Por supuesto, toda esta revolución de datos necesita un

almacenamiento adecuado en forma de **computación en la nube**.

Existen algunas **limitaciones potenciales** del big data en biomedicina. En primer lugar, está el problema de "**señal-a-ruido**" (**signal-to-noise**): datos verdaderos de la enfermedad que son sobrepasados y camuflados por datos que realmente no reflejan la enfermedad. Esto podría resultar en diagnósticos y/o tratamientos de la enfermedad por exceso o defecto. En segundo lugar, una voluminosa necesidad de datos deben ser **identificables** o al menos comprensibles o visualizables no solo por los médicos sino también por los pacientes. Tercero, algunos sanitarios pueden sentirse intimidados por los aspectos relacionados con la toma de decisiones de grandes datos cuando se combina con la inteligencia artificial y que esas máquinas inteligentes puedan **reemplazar** a algún tipo de profesional de salud. Tal vez la actitud apropiada es crear sinergias humano-ordenador que supere las capacidades de cualquiera de los dos, algo no muy diferente de cómo el GPS ayuda al conductor (aunque el automóvil sin conductor está cada vez más cerca de reemplazar al conductor). También hay una preocupación comprensible de que este cambio de paradigma provoque una **despersonalización** de la medicina, aunque lo contrario puede también ser cierto: los médicos se liberarían así de la carga del mantenimiento y el estrés de tomar la decisión médica más apropiada. Existe el problema potencial de una **capacidad de almacenamiento de datos insuficiente**, incluso en la nube. Esto podría resolverse con soluciones innovadoras como el uso del ADN como un medio digital de almacenamiento [17]. Finalmente, está la cuestión de la propiedad de los datos con información privada que pueden conducir a la identificación de pacientes (especialmente en los procesos llamados de triangulación). Este problema puede ser parcialmente mitigado por mecanismos de anonimización.

El Futuro: Una convergencia Clínico-Digital

Aunque los Estados Unidos destacan en tecnología e innovación, su aplicación se encuentra muy por detrás de los países europeos en cuanto al acceso a datos de salud, organización de la transparencia, gestión y análisis. Nuestra organización de salud y los imperativos programáticos en esta era actual deberían ir incorporando pericia en el uso de big data y análisis de datos y experiencia en inteligencia artificial para así cosechar datos médicos significativos e información que pueda estar oculta para mejorar la calidad y el resultado de la atención sanitaria [18].

En resumen, el big data y una sólida gestión de datos junto con el análisis de datos y la inteligencia artificial conducirá a un nuevo paradigma de información y conocimiento ("inteligencia médica"). Por lo tanto, nos corresponde a nosotros repensar la información médica de las tediosas formas de las tradicionales bases de datos / registros e incluso en los ensayos clínicos aleatorizados para innovar este valioso activo y crear una convergencia clínico-digital. Por lo tanto, Big Data no tiene que ver con una colección de herramientas tecnológicas y analíticas, sino más bien con una nueva transformación filosófica en nuestro ecosistema de salud en el que los datos médicos, la inteligencia artificial y la personalización de la salud está inextricablemente entrelazada con el intelecto humano y la inteligencia de la máquina.

Referencias

- [1] McAfee A, et al. Big data: the management revolution. Harv Bus Rev October 2012:60-8.
- [2] Minelli M, et al. Big data, big analytics: emerging business intelligence and analytic trends for today's businesses. Hoboken, New Jersey: JohnWiley and Sons, Inc.; 2013.
- [3] Cosentino T. Into the river: how big data, the long tail, and situated cognition are changing the world of market insights forever. LLC, Portland, Oregon: Foundational Insights; 2011
- [4] The Baseball Prospectus Team of Experts. Baseball between the numbers: why everything you know about the game is wrong. New York, New York: Prospectus Entertainment Ventures LLC/Basic Books; 2006.
- [5] Silver N. The signal and the noise: why so many predictions fail but some don't. New York, New York: The Penguin Press; 2012.
- [6] Hughes GMD. How big is big data in healthcare? From a shot in the arm blog; October 21, 2011.
- [7] Butte A (Chief, System Medicine at Stanford School of Medicine). Personal communication (February 2013).
- [8] Ashley EA, et al. Clinical evaluation incorporating a personal genome. Lancet 2010;375(9725):1525-35.
- [9] Butte A, et al. Computational translating molecular discoveries into tools for medicine: translational bioinformatics articles now featured in JAMIA. J Am Med Inform Assoc 2011;18(4):352-3.
- [10] The ENCODE. (ENCyclopedia Of DNA Elements) project. The ENCODE project consortium. Science 2004;306(5696):636-40.

- [11] Reshef DN, et al. Detecting novel associations in large data sets. Science 2011;334(6062):1518-24.
- [12] McAfee A. What's the big deal about big data? Harv Bus Rev webinar October 4, 2012.
- [13] Halevy A, et al. The unreasonable effectiveness of data. IEEE March/April, 2009;8:12.
- [14] Chang AC and Musen M. Artificial intelligence in pediatric cardiology: an innovative transformation in patient care, clinical research, and medical education. Congenit Cardiol Today, 10(11): 1-15.
- [15] Hanson CW, et al. Artificial intelligence applications in the intensive care unit. Crit Care Med 2001;29:427-35.
- [16] Hoefen R, et al. In silico cardiac risk assessment in patients with long QT syndrome. J Am Coll Cardiol 2012;60:2182-91.
- [17] Church GM, et al. Next-generation digital information storage in DNA. Science 2012;337(6102):1628-31.
- [18] Patel JL, et al. The coming of age of artificial intelligence in medicine. Artif Intell Med 2009;46(1):5-17.

(*) Este artículo es una adaptación del original: Chang AC. Big data in medicine: The upcoming artificial intelligence. Progress in Pediatric Cardiology 2016;43:91-4