PAPER 2 - BDAA FALL 2016

Carlos Sathler
Graduate Student, MS in Data Science
Indiana University at Bloomington
cssathler@gmail.com

ABSTRACT

In this paper, we explain the role of big data in health sciences, and discuss telemedicine, personalized medicine, personal monitors and privacy issues.

1. THE ROLE OF BIG DATA IN HEALTH SCIENCES

The role of big data in health sciences is to enable access to vast amounts of health data collected from multiple sources over the years, incorporate new health data originated from current technologies and aggregate this data in meaningful ways, so analytics can be used to improve several aspects of health-care and advance health sciences in general.

There are many worthy goals in health sciences and health-care, especially in the United States, which is ranked surprisingly low among other first world nations in terms of quality of care compared to cost of care. Meeker in her 2014 internet trends report [12] provides a few telling numbers. For example, the cost of our health care system was estimated in 17% of our GDP in 2014, of which an estimate 27% was waste spending. She estimated that 25% of an American family income in 2015 would be allocated to health-care, up from 18% in 2005. And medical bills are the single most prevalent cause of personal bankruptcies in the US.

According to Groves at al. [11] big data in health-care include electronic medical records, clinical trials, data from medical devices and medical images. The authors identify four data pools that need to be integrated to unleash the full potential of big data and analytics in health-care; these "primary data pools are at the heart of the big-data revolution in health-care: activity (claims) and cost data, clinical data, pharmaceutical RD data, patient behavior and sentiment data" [11, p 4]. This data comes from disparate sources such as pharmaceuticals, hospitals, laboratories physician offices and government. The volume of this data and its heterogeneity alone far exceed the capacity and capabilities of traditional structured database systems.

The opportunities for insight and discovery from analytics run against health big data have the potential to impact several areas of health sciences and health-care in particular. The Accenture's Industrial Internet Insights Report for 2015 [3] states that among many health-care organizations selected for the report, "an overwhelming majority acknowledged the critical role of analytics in driving improved clinical, financial and operational outcomes" [3, p 16]. In other words, big data and analytics have a holistic positive impact in health-care. Some examples from the report include improved patient flow, improved patient outcomes thanks to better clinical procedures, and better utilization of medical equipment.

Kaggle, a prestigious data science company runs a popular web site (kaggle.com) where data scientists use machine-learning to compete for prizes to solve real world problems. The web site offers good evidence of the potential of analytics to solve health sciences challenges. A quick online inquiry on recent Kaggle competitions [4] reveals a long list of projects in the health area. Some interesting ones are: "Cervical Cancer Screening (Help prevent cervical cancer by identifying at-risk populations)", "Ultrasound Nerve Segmentation (Identify nerve structures in ultrasound images)" and "Diabetic Retinopathy Detection (Identify signs of diabetic retinopathy in eye image)" [4, p 1].

The two examples above using medical image data were chosen to highlight the importance of that type of data in health sciences. In fact, there is a multitude of different types of image data that is relevant to the field. According to Fox et al. [10] types of medical images "include magnetic resonance imaging (MRI), digital mammography, positron emission tomography (PET), and X-ray computed tomography (CT). Together, these images amount to over 20 thousand terabytes of data" [10, p 4].

The use of genomics data in health sciences is noteworthy as well for its size and potential to help us understand diseases and provide individualized care (see section Personalized Medicine below). Thanks to the steep decline in the cost of sequencing the human genome Fox et al. [10] entertain the possibility that one day we may routinely perform genome sequencing on newborns to predict risks of diseases. The authors estimate that sequencing the genome of every new born in the US would generate 10 exabytes per year in genome data. Only through the adoption of big data technologies and analytics solutions will we be able to extract value from such volume of data.

Vinod Khosla, a venture capitalist from the Silicon Valley, predicted in 2012 "technology will replace 80% of what doctors do" [8, p 1]. He was referring to the potential that data driven approaches have to deliver improved diagnostics and prescriptions among other benefits. In 2016 his prediction still sounds like an exaggeration, but whether it will on day come true or not, it is undeniable that big data and analytics will improve many aspects of health sciences and health-care.

2. TELEMEDICINE

Telemedicine refers to the use of cloud technology to deliver medical services remotely. Since patients and providers are not collocated, telemedicine is also referred to as virtual-medicine.

Examples of telemedicine include doctor-patient exchanges via email and use of video-conferencing to deliver consultations. These practices have the potential to benefit isolated populations in remote areas of the world, with limited to no access to medical professionals. An interesting example of telemedicine comes from the TV Series House, in the episode Frozen [6]. In that episode Dr. House, somewhere in the US, has to diagnose and try to save the life of a dying scientist who is located in a basement in a research facility in the South Pole, through video-conferencing.

In its 2015 Health Report [2] inVentivHealth predicted that "the number of doctor-patient video consults in the US will nearly triple" [p 1] in 2016.

3. PERSONALIZED MEDICINE

According to Tony Hey [7] personalized medicine refers to the use of a person's genetic make-up to drive medical treatment. The person's genomic information would help determine what particular medicine should be used for a given individual, in what dosage. This approach would also allow certain drugs to be approved only for certain patients, and not for others, because they would be ineffectual or even harmful (in the form of severe undesirable side effects).

Seath Earley [9] writes of the benefits of personalized medicine to larger groups of people, stating that analysis of genomic data "could identify the most effective treatments for specific subpopulations at a more granular level than has been possible before" [9, p 8].

4. PERSONAL MONITORS

Personal monitors are devices that individuals carry with/on them to collect different type of health data. A quick search of the web revealed a significant number of different personal monitors [1][5], for example, for fitness tracking, radiation exposure, communicating location and distress, capturing vital signs (such as respiration rate and skin temperature), and tracking posture. Makers of personal monitors are manufacturing devices that can be used on the wrist as wristwatches, glasses, rings, skin patches, helmet-like contraptions, and even foot soles and textiles ("smart clothing").

Personal devices typically connect to applications on a person's smart phone or personal computer via Bluetooth technology. They allow the person to track their health data against predefined goals, for example, or it may simply display statistics and plot graphs on the different data for informational purposes. Some devices, particularly useful for providing care for the elderly, will allow nurses and other caregivers to monitor patient location. Fitness trackers like Fitbit [5] are popular personal monitors and will track physical activity such as steps, distance traveled, hours of sleep and many other personal fitness related data. Fitbit allows users to share data over the web, thus encouraging friends to create support groups as they work together and challenge each other to achieve fitness goals.

Meeker [12] reports that 75% of medical spending in the US is due to chronic conditions such as diabetes and hypertension, for example. Personal monitors have the potential to help individuals and caregivers more closely track chronic conditions; fitness trackers could help individuals achieve fitness goals that will reduce their risks of developing chronic diseases.

5. PRIVACY ISSUES

Privacy issues are a big challenge for big data and analytics in the area of health informatics. As stated before, big data has the potential to impact health sciences and health-care particularly through the aggregation of data from pharmaceuticals (clinical trials, for example), government (health programs and demographics), point of care (patient consultation data), laboratories (lab results), among other sources. Some of this data will be identifiable to individuals, by design, and only after their approval; some won't.

Wigan and Clark [13] highlight potential issues that could result from aggregating data generated from different sources, collected for different purposes. De-identification of health data is imperative if we are to protect an individual's privacy over their health information, while we try to use it to advance the health sciences, improve the quality of our health-care system, and reduce costs of care giving.

6. REFERENCES

[1] 10 Wearable Health Tech Devices To Watch - InformationWeek. http://www.informationweek.com/mobile/10-wearable-health-tech-devices-to-watch/d/d-id/1107148?page$_n umber =$ 1.$(Accessedon$09/02/2016).

[2] 2015 health trends. http://www.slideshare.net/GSW_Worldwide/2015-healthtrends. (Accessed on 08/28/2016).

[3] Accenture Industrial Internet Insight Reports for 2015. https://www.accenture.com/us-en/_acnmedia/Accenture/nextgen/reassembling-industry/pdf/AccentureIndustrialInternetChanging-CompetitiveLandscapeIndustries.pdf. (Accessed on 08/28/2016).

[4] Competitions | Kaggle. https://www.kaggle.com/competitions. (Accessed on 09/02/2016).

[5] Fitbit Surge. https://www.fitbit.com/shop/surge. (Accessed on 09/02/2016).

[6] "House" Frozen (TV Episode 2008) - IMDb. http://www.imdb.com/title/tt1168087/. (Accessed on 08/31/2016).

[7] Semantic Computing and Smart Cyberinfrastructure. http://www.ieee-icsc.org/ICSC2010/Tony%20Hey%20-%2020100923.pdf. (Accessed on 08/28/2016).

[8] Technology will replace 80% of what doctors do. http://fortune.com/2012/12/04/technology-will-replace-80-of-what-doctors-do/. (Accessed on 09/02/2016).

[9] S. Earley. The promise of healthcare analytics. *IT Professional*, 17(2):7–9, 2015.

[10] G. Fox, T. Hey, and A. Trefethen. Where does all the data come from. http://grids.ucs.indiana.edu/ptliupages/publications/where does all the data come from v7.pdf. (Accessed on 08/30/2016).

[11] P. Groves, B. Kayyali, D. Knott, and S. Van Kuiken. The 'big data' revolution in healthcare. *McKinsey Quarterly*, 2, 2013.

[12] Mary Meeker. 2014 internet trends. Web Page, June 2014. [accessed 2016 October 27] Available from: http://www.kpcb.com/blog/2014-internet-trends.

[13] M. R. Wigan and R. Clarke. Big data's big unintended consequences. *Computer*, 46(6):46–53, 2013.