BIG DATA IN HEALTH CARE DELIVERY

Kingsley Effah
Jiangsu University
Zhenjiang, China

Cai Tao
Jiangsu University
Zhenjiang, China

Xue Rui
Jiangsu University
Zhenjiang, China

ABSTRACT — Health Care is one of the biggest social and economic issues in many countries and the world as a whole. Big Data in healthcare refers to an electronic data which can unify all patient health related data to get 360-degree view of patient well-being in other to scrutinize and envisage outcomes. Data in the healthcare environment keeps on increasing beyond dealing ability of the healthcare delivery organization and is expected to keep on increasing significantly in the coming years. Big data can help improve health care practice, new drug development and management of healthcare delivery system and it financing process. Big data offer a lot of benefits to health care delivery such as early disease discovery and prevention, monitoring and for better health care delivery and efficacy.

This paper introduces big data analytics concept in healthcare delivery management and it characteristics, healthcare data and some key subjects of big data. The advantages and challenges of big data in healthcare are also discuss.

Keywords: Analytics, Big Data, Database, Data Mining, Diagnosis, Fraud Detection, Hadoop, National Health Insurance, Personalized Patient Care.

INTRODUCTION

Big Data mostly refers to data that exceeds the processing capacity of conversional database system in terms of storage, processing, computing capacity and data analysis techniques. The data is big in volume, move fast and also difficult to manage using the traditional database architectures. The healthcare sector from the past has generated large amount of data, in the areas of record keeping, compliance & regulatory requirements, and patient care.

Big data in healthcare delivery generally refers to the collection of electronic health data sets that are huge in volume and complex to process using the conventional database management tools or the traditional approach of managing data and their applications. Data-intensive, parallel and distributed methodologies are usually employed, such as the MapReduce programming paradigm (e.g. Apache Hadoop). Data in healthcare includes patient care data such as physician notes, Laboratory reports, case history, and social media post, X-Ray report, and list of employee in a particular healthcare delivery point.

910

HEALTHCARE AND BIG DATA CHARACTERISTICS (5 V'S)

The term Big Data is widely used recently in order to explain data in the physical, biological and the cyber world. There is no single formal definition of Big Data. However, most of the definitions seems to coalesce around the 5 V's: Volume, Velocity, Variety, Veracity and Value. . By understanding the 5V's of big data, we can then apply their power in research and real problem solving.
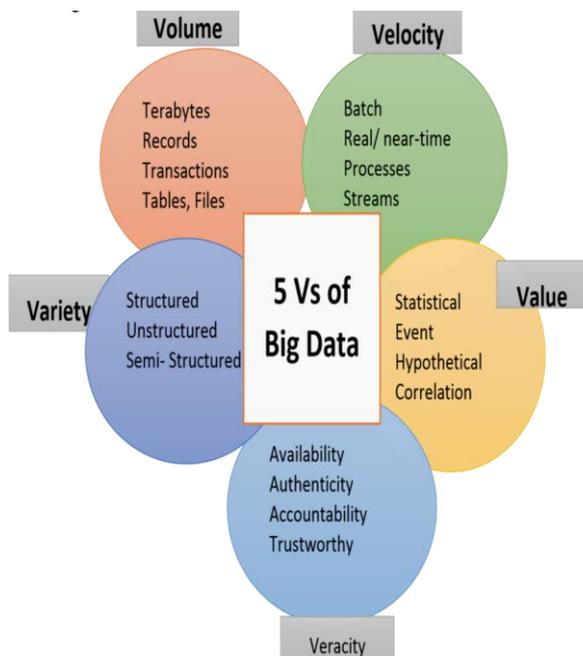

FIGURE: 1. the 5V's of Big Data

Volume: The quantity (terabytes, petabytes, etc…) of healthcare data been generated to be analyzed. These include healthcare data such as: personal medical records, radiology images, biometric sensor reading, and genomic. Healthcare system can now have the potential to manage and analyse complex data. Healthcare data keeps on increasing as the day goes by.

Variety: Healthcare data can be classified as structured, unstructured and semi-structured. Structured healthcare data such as clinical data, are easy to manipulate, store and analyse by machine. Most of healthcare data, such as office medical records, doctor notes, paper prescription, and radiograph films are unstructured or semi-structured.

Velocity: Velocity in healthcare refers to how fast healthcare data is been generated and processed for effective and efficient delivery of healthcare. Most data in the healthcare sector has been traditionally static; paper file, x-ray films, and script. In medical situations, constant real time data (operating room monitors for anesthesia, beside heart monitor, etc..,) can mean the difference between life and death.

Veracity: This characteristic of big data is the goal to make sure big data analytics and results are prone to error free and trustworthy. In healthcare delivery data quality issues are of great concern for two purposes: life or death decision depend on having the precise information, and the quality of healthcare data, especially unstructured data are highly variable and too often incorrect (Inaccurate transformation of poor handwriting on prescription).

Value: In order to collect and extract maximum true value from big data we work

911

with, then we must look into data storage investment, because value of data depends on quality of governance strategy and mechanism. For example, storing clinical data for new diseases on unreliable storage can save money today, but may affect data tomorrow. And also, another point to consider is that some data in time of collection have different value to risk, but this risk can developed in time.

HEALTHCARE DATA

Big data in healthcare refers to the electronic data sets so huge and complex to store and analyze with traditional data processing applications or difficult to process using common database management tools. For big data healthcare system, the Hadoop MapReduce framework is uniquely capable of storing a wide range of healthcare data types and offers high scalability, reliability and availability than traditional database management system. Healthcare data can be categorized as follows:

Genomic Data: it refers to genotyping, gene expression and DNA sequence (Chen et al., 2012; Priyanka and Kulennavar, 2014).

Clinical Data and Notes: About 80 to 85 percent of these type of data are unstructured documents, clinical or transcribed notes and images (e.g., post-op notes, diagnostic testing reports, patient discharge summaries, medical images such as radiological image and x-ray images.)

Health Publication and Clinical Reference Data: Text-book publications (journals, clinical research and medical reference material) and clinical text-based reference guidelines and health product (e.g., drug information) data (Miller, 2012; Priyanka and Kulennavar, 2014).

Behavior Data and Patient Sentiment Data: Mobility sensor data or streamed data (data in motion, e.g., electroencephalography data). They are from regular medical monitoring and home monitoring, telehealth, sensor-based wireless and smart devices (Shrestha, 2014).

Administrative, Business and External Data: Data ranging from insurance claims and related financial data, billing and scheduling. Biometric data such as: fingerprints, handwriting, iris scans, and etc.

Other Important Data: The content from portal or patient's health records messaging such as e-mail between the patient and the provider; device data, adverse event and patient feedback, etc.

IMPACT OF BIG DATA ANALYTICS ON HEALTHCARE DELIVERY

Good living: Patient can build value by taking an active role in their own treatment, including disease prevention. The right living focuses on encouraging patient to make lifestyle decisions that will help maintain healthy living, such as proper nutrition and exercising.

Innovation: It includes the identification of new treatment and approaches in healthcare delivery across all aspects of the system and by improving the innovation engines. They could also use the data to find opportunities to improve clinical trials and traditional treatments, including those for births and inpatient surgeries.

Right care: This is to make sure that patient have access to timely and appropriate treatment available. Right care requires a coordinated approach across settings and providers, all health facilities should have the same information and work towards the same goal and objectives to elude duplication of effort and suboptimal policies.

Value for Money: To fulfil the goals of value for money, providers and payers will continuously enhance healthcare delivery value while improving its quality. This could involve multiple approaches for ensuring cost-effective of healthcare delivery system, by eliminating scam, mishandling and waste in the system.

ADVANTAGES TO HEALTHCARE DELIVERY.

Data is an important tool to advance population healthcare and developing new types of personalized healthcare delivery. Big data is also helping consumers acquire more reliable and timely information about cost and quality care. Applying progressive analytics to healthcare data can help detecting diseases at earlier stages, detecting healthcare abuse and fraud faster, and reducing cost.

Benefits to Clinical Operation: Big data can benefit in defining more clinically related and cost effective ways to healthcare delivery and patient treatment. It could also help in reduction of cost in healthcare delivery by analysing disease patterns and track disease outbreaks and transmission to enhance public healthcare delivery and better response.

Benefits to Research & Development: Collecting health data from a diverse sources and analysing them together can help advance about new diseases and treatments. Research & development contribute to new techniques and tools, such as the techniques by Facebook, Google and twitter that explain what patient find about their health system. It improve predictive models for producing more devices and therapies for the market and also identify related on symptoms and determine adverse effects before healthcare products get into the market. Also having a centralized data for healthcare system can also aid in learning by providing precise and on time information for medical students for educational purpose and research work.

Benefits to Fraud Detection and Prevention: Using advanced analyses systems implementation can help identify, predict and minimize fraud by preventing of inaccurate and wrong health claims submitted by healthcare providers. The predictive modelling of big data can be used by healthcare organization for

fraud prevention in order to promote or for fill value for money in healthcare delivery.

Benefits to Personalized Patient Care: currently, healthcare delivery is shifting from a disease centered model towards a patient centered model. In disease centered model, decisions making are centered on the medical know-how and data from clinical evidence and numerous tests. Big data is vital in patient centered model in healthcare deliver by creating a personalized disease risk and management plan and awareness plan for an individual.

CHALLENGES OF BIG DATA IN HEALTHCARE DELIVERY

The 5 Vs of big data in healthcare comes with challenges in supporting the very key functions necessary for data storage, curation, retrieval, search and visualization. Some of the criteria for evaluating big data platform in healthcare delivery include security, availability, quality assurance, continuity, scalability, privacy and ease of use.

Lack of Centralized Database: The sources of data from the healthcare organizations like the hospital, pharmacies, medical centres, etc. are in different format and also have their data on different systems and settings. In order to have access to this big amount of data from all these healthcare providers, there must be a common data warehouse (centralized database) in order to get standardized information for effective management and easy access.

Because of competitive advantage many healthcare providers are reluctant to share data and however, building such a system comes with additional cost.

Data Privacy and Security: Protecting patient privacy and it health records from hackers has become a challenging factor in big data. There have been some security measures in recent years to ensure data privacy and security in the big data environment but it is almost impossible to guarantee hundred percent data security. Is sometimes difficult to determine the balance between protecting the integrity of patient's information and usability of data.

Cost: Investing in big data comes with huge cost for organizations to hire employee (data scientists) with high level of expertise and technical knowledge in using data mining and big data analytics tools. Healthcare organizations must also investment in infrastructure, policies and standards of practice to ensure effective running of the system which also come with a huge cost.
Lack of Data Quality: Due to the fact that majority of data collected in the healthcare environment are unstructured (e.g., diagnostic testing reports, patient discharge summaries, and x-ray images) make it challenging to aggregate and analyze them to extract potential useful information for decision making. Due to the serious constrains regarding the quality of collected data, results are prone to errors and disparities.

CONCLUSION

Big data in healthcare delivery is based on data acquire from the process of diagnosis through to treatment of each patient. Big data have the potential to improve healthcare delivery by changing healthcare delivery from reporting to predicting results at earlier stages. Big data offers a lot of benefits which include disease prevention, fraud detection, value-for-money healthcare deliver, research and development, improve medicine and also have the potential of reducing medical errors.   Challenges in big data such as data integration, data visualization and information security and privacy will be overcome with advance in information technology and other disciplines and support from agencies and government in policy making. This paper presented a general vision of big data for acquiring, analysing and managing data in healthcare environment in different forms and from different sources. Big data application in healthcare delivery is at an emerging stage of development, but with rapid advance in research, tools and technologies can accelerate their emerging process.

REFERENCES:

[1] Schmitt, C., M. Shoffner and P. Owen, 2013. Security and privacy in the era of big data: The SMW, a technological solution to the challenge of data leakage. RENCI, University of North Carolina at Chapel Hill.

[2] Bottles, K. and E. Begoli, 2014. Understanding the pros and cons of big data analytics.

Physician Exec., 40: 6-12.

[3] Bulletin Board, 2014. Industry assesses potential, challenges of big data. J. AHIMA.

Chawla, N.V. and D.A. Davis, 2013. Bringing big data to personalized healthcare: A patient entered framework. J. General Internal Med., 28: S660-S665. DOI: 10.1007/s116060132455-8

[4] Chen, H.C., R.H.L. Chiang and V.C. Storey, 2012. Business intelligence and analytics: From big data to big impact. MIS Q., 36: 1165-1188.

[5] Demchenko, Y., P. Grosso and C. de Laat, 2013. Addressing big data issues in scientific data infrastructure. Proceedings of the International Conference on Collaboration Technologies and Systems, May 20-24, IEEE Xplore Press, San Diego, CA, USA, pp: 48-55.

[6] Dumbill, E., 2013. Making sense of Big Data. Big Data.

[7] Eaton, C., D. Deroos, T. Deutsch, G. Lapis and P. Zikopoulos, 2012. Understanding big data.

McGraw-Hill Companies.

[8] Erdman, A.G. and D.F. Keefe, 2013. Grand challenge: Applying regulatory science and big data to improve medical device innovation. IEEE Ttrans. Biomed.

[9] Bellini, P., M. d. Claudio, P. Nesi and N. Rauch, 2013. Tassonomy and Review of Big Data Solutions Navigation. In: Big Data Computing, Akerkar, R. (Ed.),   Chapman and Hall/CRC, pp: 57-101.

[10] Halamka, J.D., 2014. Early experiences with big data at an academic medical center.

Health

Affairs,33:1132-1138.DOI:10.1377/hlthaff.20 14.0031

[11] Helm-Murtagh, S.C., 2014. Use of Big Data by Blue Cross and Blue Shield of North Carolina.

NCMJ, 75: 195-197.

[12] Hsieh, J.C., A.H. Li and C.C. Yang, 2013. Mobile, cloud and big data computing: Contributions, challenges and new directions in telecardiology. Int. J. Environ.

[13] Jagadish, H.V., J. Gehrke, A. Labrinidis, Y. Papakonstantinou and J.M. Patel et al., 2014. Big data and its technical challenges. Communications

[14] Lee, J.H. and E.J. Lee, 2014. Computer-aided diagnosis sensor and system of breast sonography:

[15] Manyika, J., M. Chui, B. Brown, J. Buhin and R. Dobbs et al., 2011. Big data: The next frontier for innovation, competition and productivity. USA: McKinsey Global Institute.

[16] Megahed, F.M. and L.A. Jones-Farmer, 2013. A Statistical Process Monitoring Perspective on "Big Data", Frontiers in Statistical Quality Control, 11th Edn. Springer, New York.

[17] Miller, K., 2012. Big data analytics in biomedical research. Biomedical Computation Review. Minelli, M., M. Chambers and A. Dhiraj, 2013. Big Data, Big Analytics: Emerging Business

[18] Intelligence and Analytic Trends for Today's Business. 1st Edn. John Wiley & Sons, Inc. Hoboken, New Jersey, USA, ISBN-10: 111814760X, pp: 187.

[19] Ngufor, C. and J. Wojtusiak, 2013. Learning from largescale distributed health data: An approximate logistic regression approach. Proceedings of the 30th International Conference on Machine Learning, Atlanta, Georgia, USA, JMLR: W&CP, pp: 1-8.

[20] O'Reilly Radar Team, 2012. Planning for big data. O'Reilly.

[21] Priyanka, K. and N. Kulennavar, 2014. A survey on big data analytics in health care. Int. J. Comput. Sci. Inform. Technologies, 5: 5865-5868.

[22] Raghupathi, W. and V. Raghupathi, 2014. Big data analytics in healthcare: Promise and potential.

[23] Rajpathak, T. and A. Narsingpurkar, 2013. Managing knowledge from big data analytics in product development. White Paper, Tata Consultancy Services.

[24] Russom, P., 2011. Big data analytics. TDWI Best Practices Report, Fourth Quarter.

[25] L. Aniket Bochare "Heterogeneous Data Integration for Clinical Decision Support System" 2011.

[26] K. R. Ghani, K. Zheng, J. T. Wei, and C. P. Friedman, "Harnessing big data for health care and research: are urologists ready?" Eur. Urol., vol. 66, no. 6, pp. 975–7, Dec. 2016.

[27] D. Lazer, R. Kennedy, G. King, and A. Vespignani, "The Parable of Google Flu: Traps in Big Data Analysis," *Science (80-.)*, vol. 343, pp. 1203–1205, 2014.

First Author: Kingsley Effah

School: Jiangsu University, China

Department: School of Computer Science and Communication Engineering

Major: Master in Computer Science and Technology (Big Data).

Second Author: Cai Tao

School: Jiangsu University, China

Department: School of Computer Science and Communication Engineering

Associate Professor, Jiangsu University-School of Computer Science and Communication Engineering**,**

Third Author: Xue Rui

School: Jiangsu University, China

Department: School of Computer Science and Communication Engineering

Major: Master in Computer Science and Technology

917