

A BIG DATA REVOLUTION IN HEALTH CARE SECTOR: OPPORTUNITIES, CHALLENGES AND TECHNOLOGICAL ADVANCEMENTS

Sanskriti Patel and Atul Patel

Faculty of Computer Science & Applications, CHARUSAT, Changa, India

ABSTRACT

Health care sector grows tremendously in last few decades. The health care sector has generated huge amounts of data that has huge volume, enormous velocity and vast variety. Also it comes from a variety of new sources as hospitals are now tend to implemented electronic health record (EHR) systems. These sources have strained the existing capabilities of existing conventional relational database management systems. In such scenario, Big data solutions offer to harness these massive, heterogeneous and complex data sets to obtain more meaningful and knowledgeable information.

This paper basically studies the impact of implementing the big data solutions on the healthcare sector, the potential opportunities, challenges and available platform and tools to implement Big data analytics in health care sector.

KEYWORDS

Big Data, Health Care, Big Data Analytics

1. INTRODUCTION

The health care sector grows rapidly in last 30 years. The healthcare industry historically has generated large amounts of data, driven by record keeping, compliance & regulatory requirements and patient care. While most data is stored in hard copy form, the current trend is towards the rapid digitization of these large amounts of data. There are different types of data sources which generates these enormous amounts of data. Big data in healthcare refers to electronic health care records (EHR) that is quite large and complex that they are difficult to manage with traditional software and/or hardware. Also, they are not easily managed with traditional or common data management tools and methods. Using the technologies that able to deal with such “Big Data” will offer many potential opportunities to the healthcare sector.

This research paper aims to deal with the main opportunities and challenges of the big data and its analytics in healthcare. It also discusses the current technological platform and tools that can help to utilize big data effectively. Section 1 of the paper gives brief introduction about Big data and its characteristics. Section 2 provides information on health care sector and Big data. Section 3 provides an insight of Big data analytics. Section 4 demonstrates the main opportunities of Big data in health care sector, while section 5 discusses the major challenges and threats of Big data

implementation in health care sector. The available platform and tools for Big data implementation is presented in Section 6. Finally, the paper's conclusion is in Section 7.

1.1 Big Data: Background and its Sources

Big data is a term that is used to describe large volume of data. Data may in form of structured or unstructured. The analytics of Big data leads to any organization towards better decision making and strategic steps. Giant companies in sectors like retail, manufacture and government agencies are using Big data to meet their business and strategic objectives. The Big data analytics also plays a vital role for small and medium size industries to capitalize their business.

Industry analyst Doug Laney originally coined the concept of Big data while referring to the challenge of data management [8]. According to that, there are three important dimensions of the Big data concept illustrated below [5].

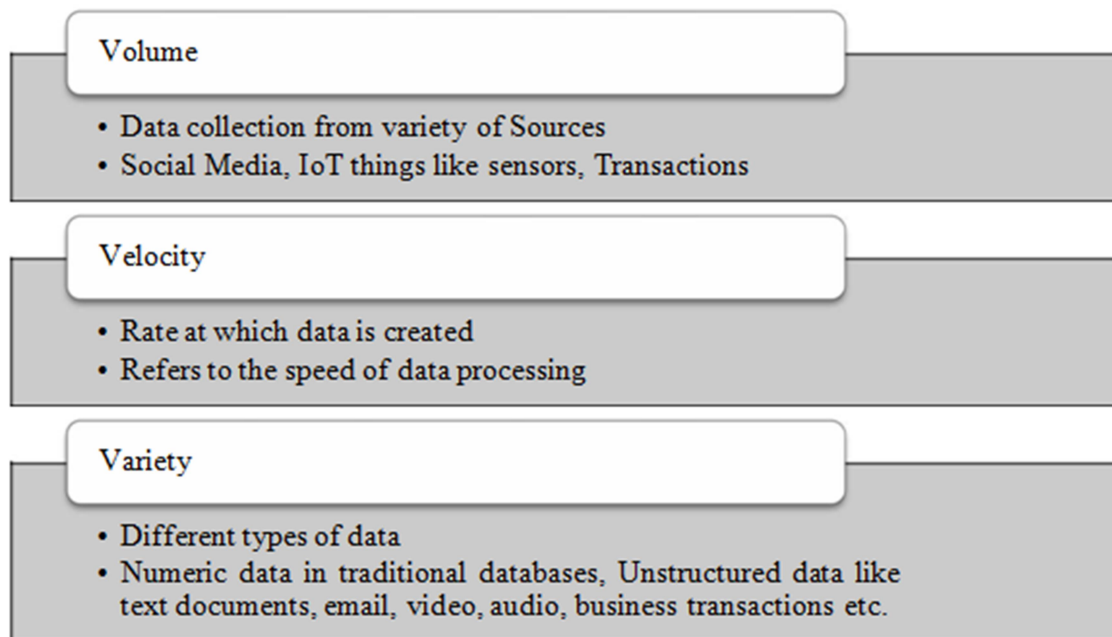


Figure 1. Three Vs of Big Data

Today, many organizations are gathering, storing, and analyzing huge amounts of data. These data is known as a Big data as it has Volume, Velocity and Variety. Gartner [2012] predicts that by 2015 the need to support big data will create 4.4 million IT jobs globally, with 1.9 million of them in the U.S.[9]

1.2 BIG DATA SOURCES

There is variety of sources from where the Big Data is generated. Social Media sources such as Facebook, Instagram, Twitter generates terabytes of data on every day. Machines such as desktop computer, laptop generates tremendous amount of data. Geospatial data is generated by cell phones and even from satellites. The IoT (Internet of Things) devices like sensors, pocket computers are also generating a massive amount of data. Data is also generated as an output of research projects like Large Hadron Collider (LHC) at CERN in Switzerland and France output an enormous amount of data - over 200 petabytess [15].

Several sectors are benefited from the Big Data analytics like The Financial Services Industry, The Automotive Industry, Supply Chain, Logistics, and Industrial Engineering, Retail, Health care, Entertainment etc. Combining Big Data with Analytics leads to any organization towards many tasks like determining root causes of collapses, issues and defects in near-real time, generating coupons based on the customer's buying habits at the point of sale, recalculating entire risk portfolios in minutes, detecting fraudulent behavior before it affects your organization etc. [17]

2. HEALTH CARE AND BIG DATA

An information and communications technology (ICT) is playing a vital role in improving health care for individuals and communities. It helps to improve health system efficiencies and prevent medical errors. With an invent of new and efficient mechanisms for storing and accessing information, ICT helps to serve a society in a better way. ICT powered health mechanisms are often known as eHealth.

One of the characteristic that health care sector possesses is its data richness. With the development in diagnostic and treatment, health care sector evolved so quickly in last few decades. There are many sources in this sector from where the data is generated. These data is undoubtedly in the form of Big Data. The data came from many sources and categorized as follows:

1. Web and social media data: Data captured from Facebook, Twitter, LinkedIn, blogs, and the like. It can also include health plan websites, smartphone apps etc. [14]
2. Machine-to-machine (M2M) device generated data: readings from remote sensors, meters, and other devices [6].
3. Biometric data: Data may in form of retinal scans, x-ray images, finger prints, genetics, handwriting, other medical images, blood pressure and other similar types of data [14].
4. Human-generated data: In the form of unstructured and semi-structured data. Some of the examples are EMRs, Doctor's notes and paper documents [14].

Genomic Data: data in the form of DNA sequence [2].

3. BIG DATA ANALYTICS

Big Data analytics is the process of exploring huge data sets that may contain a variety of data types to reveal hidden patterns, unknown correlations, market trends, customer preferences and other useful business information [1]. Big data analytics has emerged from two distinct concepts: big data and analytics. Big Data analytics in Healthcare is fundamentally a set of methodologies, procedures, frameworks and technologies which are used to transform raw data into meaningful as well as useful information. These set of information are used to make decision making tasks more effective whether they are strategic, tactical & operational. The following figure 1[10] depicted the key components playing a role in Big Data analytics for health care sector [10].

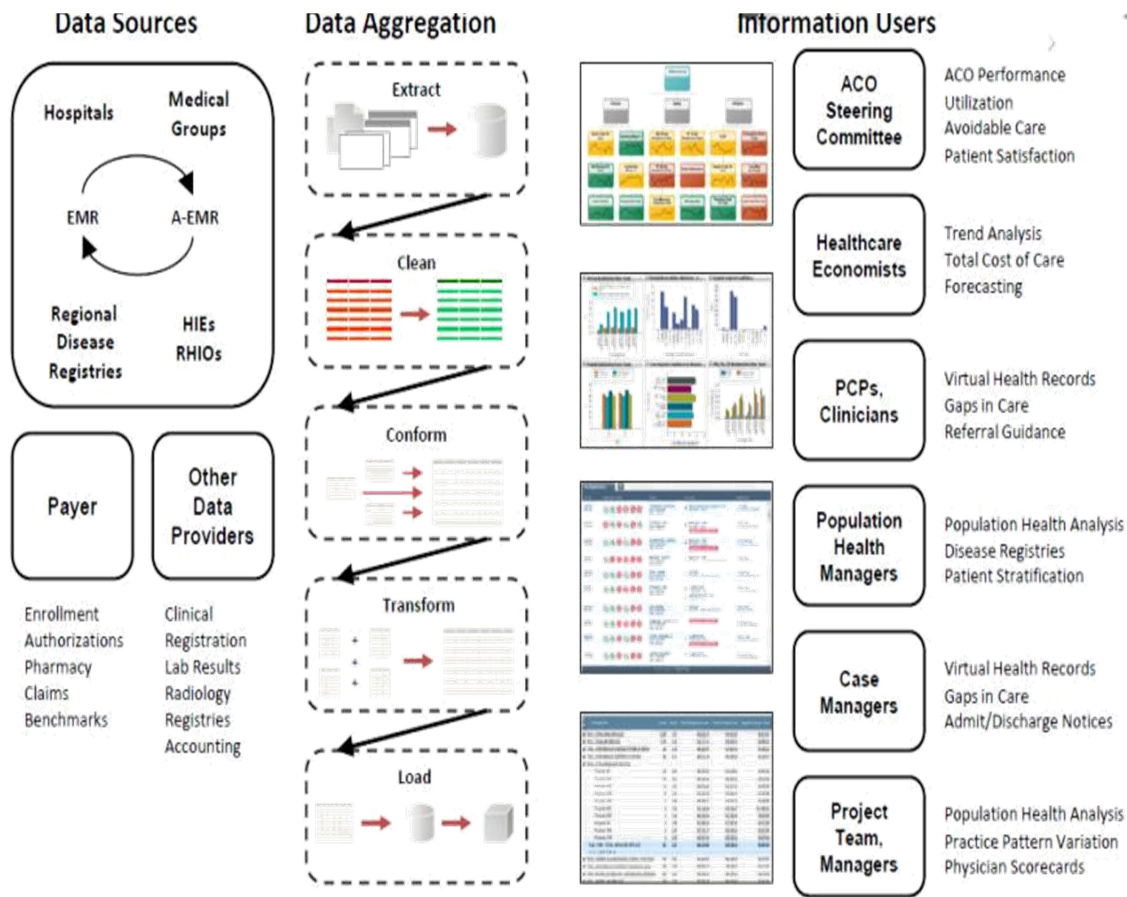


Figure 1. Big Data Analytics in Health Care: Key Components

As per the figure 1, data produces from variety of sources like hospitals, medical groups, payers or other data providers. These data first needed to be aggregated. Moreover, many processes like extraction, cleaning, conformation, transformation and loading are executed on data during this phase. Finally, some meaningful and useful information are generated which can be used by variety of users and purpose as shown in figure 1.

4. OPPORTUNITIES OF BIG DATA IN HEALTHCARE

This section discusses the various opportunities of Big data in health care.

Decreasing Healthcare Costs to Get Financial Profit

Big data can help decrease the cost of providing medical treatment in many ways. Moreover, analysis on data gives insight to health care providers to determine populations at risk for illness. By doing so, proactive steps can be taken initially. Data and its analytics are easier than ever to share. Big data can more accurately pinpoint where education and prevention is needed to produce healthier populations at lower costs. Treatment is more evidence based using Big Data analytics[4].

Promotes Research and Innovation

By analytics on data, the current state of health of patients provides insight to them to take more ownership of their healthcare. The information sharing mechanism increases productivity and reducing overlapping of data. By thus, it is enhancing the coordination of care. [4].

Personalized Medicine

In past few years, it is possible to predict the lifestyle diseases through genetic blue prints. Big data will further personalize medicine by determining the tests and treatments needed for each patient. The provision of earlier treatment can reduce the health costs and can eliminate the risk of chronic diseases [4].

Strengthen the Preventive Care

Prevention is always better than cure. Following this thumb of rule, with the advent of Big Data analytics, it is easy to capture, analyze and compare patient symptoms earlier to offer a preventive care in a better way.

Virtual Care and Wearable Health Care Technologies

Technology is helping providers make virtual care initiatives that increase quality of care and provide patients with more access [3].

Health Trend Analysis

By using different analytical approaches including data mining and text mining techniques, health trend analysis and comprehensive patient management is more easy using Big Data Analytics [7]

Identification and Tracking of Patients

The identification and tracking of patients with type 2 diabetes is discussed in recent article [6]. The author suggests to use a two-step process to identify subsets of patients that have similar clinical indications and care patterns. In a first step, patients are divided into groups based on the primary diagnosis. Then after, a statistical clustering method is applied to further divide the subsets. This method uses readily available administrative datasets. Also, patients must be tracked longitudinally to determine the patterns for treatment. Therefore, the method is applicable in scenarios where patient data is available over time and across providers [16].

Studying Drug Efficacy

Electronic health record (EHR) data may also be used to study drug efficacy. Researchers at the University of Pennsylvania School of Medicine [15] compared the results of randomized controlled trials versus using an EMR to compare cardiovascular outcomes. It has been observed that the cost of randomized controlled trials is much higher than the cost of using readily available EHR data to compare treatment modalities [16].

5. CHALLENGES OF BIG DATA IN HEALTH CARE

Some of the challenges which make Big Data analytics difficult to use in health care are discussed below.

Protecting the Patient's Privacy

One of the significant challenges in leveraging health care's big data to its full extent is policies that protect the privacy of patient's data. Many laws protect the patient's data and not reveal the patient's identity that makes the big data analytics difficult.

On the contrary, sometimes health care providers are themselves are reluctant to share data because of market competition. A physician many not want their competitors to know exactly how many and which types of procedures they performed and where. Also, the demographics of hospitals provide one hospital a financial advantage over another. Some of the datasets are publicly available but these data sources are typically historical data or limited to government payers[16].

Data Aggregation

In health care sector, the data is in unstructured form. These unstructured data is in the form of images, graphs, notes of doctor's etc. Apart from this, the nature of structured data is mostly heterogeneous. These may lead a huge problem at the time of aggregation of these data. Natural language processing and free-text software could solve this problem at some extent but it is in its initial stage.

Cost Incurred for Establishment of Big Data Architecture

To have a benefit through Big Data analytics, it requires organization level management and analysis as well as a large-scale investment.

Requirement of Expert Knowledge

Big Data systems require data scientists with specialized experience to support design, implementation, and continued use. The McKinsey Global Institute estimates that there will be a more than 100,000 person shortage through 2020. It means that mean 50–60% of data scientist positions may go vacant. Data scientists need highly technical skill sets. They must possess soft skills such as communication, collaboration, leadership, creativity and more [11].

Security Concern

Health data is a very much personal data. Patients expected extra privacy protection if they are going to fully participate in Big Data analytics projects. In such types of projects, users should be authorized at different levels and time periods. These will prevent unauthorized access to medical records is nearly impossible [12].

6. TECHNOLOGY SUPPORT FOR BIG DATA ANALYTICS IN HEALTH CARE

There are varieties of platforms and tools are available for Big Data analytics in healthcare. Some of these are mentioned in the table 1.

Table 1. Platforms and Tools for Big Data Analytics

| | |
|---------------------------|---|
| Cloud Storage | Cloud storage uses a network of remote servers. These servers are hosted on the Internet to store, manage, and process data. There are many vendors that provide cloud storage. For example Google Cloud Storage is a key part of storing and working with Big Data on Google Cloud Platform. For Bigquery and Hadoop, using a Google Cloud Storage bucket is optional but recommended. |
| Column oriented databases | Column-oriented databases basically stores data sets as segments of columns of data rather than as rows of data. It allows huge data compression and very fast query times. |
| NoSQL databases | In relational databases tabular relations are used while a NoSQL (Not only SQL) database provides a different method for storage and retrieval of data. It focuses on storage and retrieval of huge volumes of semi-structured, unstructured or even structured data. |
| Hadoop System | Hadoop is so far the most popular implementation of MapReduce methodology. It is an entirely open source platform for handling Big Data. |
| Hive | Hive is a runtime Hadoop support architecture that leverages Structure Query Language (SQL) with the Hadoop platform. |
| PIG | PIG consists of a "Perl-like" language. Instead of a "SQL-like" language, it allows for query execution over data stored on a Hadoop cluster. |
| Cassandra | Cassandra is also a distributed database system. It is designated as a top-level project modelled to handle big data distributed across many utility servers. |

7. CONCLUSION

We may consider Big data as a latest evolution in the field of decision support data management systems. On the other side, the digitalization in health care sector is in peak. As we discussed in the paper, there are several opportunities for Big data in health care sector. Meanwhile, the technological advancement is rapidly going on towards the implementation of Big data analytics. In near future, there will be widespread implementation of big data analytics across the health care organization and the healthcare industry. The Big data solutions could definitely save millions of life and improve patient services.

REFERENCES

- [1] Big data analytics definition, retrieved from <http://searchbusinessanalytics.techtarget.com/definition/big-data-analytics>
- [2] Big Data Analytics for Healthcare, retrieved from <https://www.siam.org/meetings/sdm13/sun.pdf>
- [3] Big Data in health care, retrieved from http://www.sas.com/en_us/insights/articles/big-data/big-data-in-healthcare.html

- [4] Big Data Offers Big Opportunities in Healthcare, retrieved from <http://www.villanovau.com/resources/bi/big-data-healthcare-opportunities/#.VnfRrArZ95kg>
- [5] Big Data, retrieved from http://www.sas.com/en_us/insights/big-data/what-is-big-data.html on December 20, 2015
- [6] Bradley P. Implications of big data analytics on population health management. *Big Data*. 2013;1(3):152–159
- [7] C. Chai, "Big Data in Healthcare", retrieved from http://ihealthtran.com/pdf/Big%20Data%20in%20Healthcare_Chai.pdf
- [8] Doug Laney, Application Delivery Strategies, retrieved from <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf> on December 20, 2015
- [9] Hugh J. Watson, Tutorial: Big Data Analytics: Concepts, Technologies, and Applications, *Communications of the Association for Information Systems*, Volume 34, Article 65, pp. 1247-1268, April 2014
- [10] Informatics or Analytics? Understanding Digital Health Use Cases, retrieved from <http://practicalanalytics.co/2013/07/15/informatics-or-analytics-understanding-healthcare-provider-use-cases/>, HIMSS 2013
- [11] Overcoming Healthcare Big Data Challenges, retrieved from <http://www.mckesson.com/healthcare-analytics/healthcare-big-data-challenges/#footNote>
- [12] Smart Use of Big Data: The Key to the Future, retrieved from <http://www.healthcare.siemens.com/magazine/mso-big-data-and-healthcare-2.html>
- [13] Tannen RL, Weiner MG, Xie D. Use of primary care electronic medical record database in drug efficacy research on cardiovascular outcomes: comparison of database and randomised controlled trial findings, *BMJ*. 2009;338:b81
- [14] Transforming Health Care through Big Data Strategies for leveraging big data in the health care industry, retrieved from <http://ihealthtran.com/wordpress/2013/03/ih%20releases-big-data-research-report-download-today/>
- [15] Using Open Source to Distribute Big Data from the Large Hadron Collider, retrieved from <https://www.linux.com/news/enterprise/networking/873403-using-open-source-to-distribute-big-data-from-the-large-hadron-collider> on December 15, 2015
- [16] White SE, A review of big data in health care: challenges and opportunities, *Open Access Bioinformatics* 2014, 6:13-18, 2014
- [17] Why Is Big Data Important?, retrieved from http://www.sas.com/en_us/insights/big-data/what-is-big-data.html on December 15, 2015