

Healthcare Analytics in the Modern Era: A Survey

Mohammad Zunnun Khan¹, Mohd Shoaib²

Computer Science and Engineering¹, Zakir Husain College of Engineering and Technology²,

Integral University¹, Aligarh Muslim University²,

Lucknow¹, Aligarh²

India.

khanmzkhan@gmail.com¹, md.shoaiibs@gmail.com²

Abstract- Now we are living in the era of Big Data in information technology (IT) related fields. Health Care sector is one of the biggest effected area by information technology, and the size of data in this filed grows rapidly due to involvement of IT. The volume, type and intensity of Health Care Data expands very drastically. These days Hospitals are uploading their records online it is also a very big issue for this. So it is biggest task to analyze, manage and compute this huge amount of data received from various source on day today basis. The traditional Database Schema (Relational Schema) is not capable of doing this. Big Data offers different types of solution over these situations, it help us to manages heterogeneous type of data, different complex set of data, and to predict more useful and knowledge oriented solutions.

With the help of big data analytics technique in Health care, we can focuses on offering insights into various issues of hospital management such as patient records on global basis , costs comparison, diagnoses techniques an result efficiency. Big data in heath care also help s to offer insights on both the macro and micro level. Heath care using big data can also be important as it can also be used in easy prediction of diseases outcome suggestive measures, success and failure ratio, post diagnostics measures, etc. Various countries such as Malaysia have shifted their focus towards healthcare industry on Big Data as one of the important data storage and base. Yet there are many challenges in implementing big data especially in health care, such as security and the privacy related to patient record, different standard and governance polices etc. Internal management of data is also a tedious task like classification data, data integration, incorporating technologies.

On the basis of above discussion we can say that before implementing Big Data in Health Care, must overcome these challenges. This study specifically emphasises on various challenges, prospects, tool sets and its impact on health care industry. This study also emphasis on various mitigation approaches for the data management in healthcare.

Keywords- Big Data, Healthcare, Data analytics

1. INTRODUCTION

Healthcare has grown rapidly in the past three decade; by this growth the biggest challenge is to manage the million and trillion of data for each and every individual on minute by minute basis. It is a hectic practice to manage data in old and traditional manners. Keeping updated record and produce a traceable record for the outer world is a very difficult exercise. The sources that generate data are different in working and but to get the optimized results and smart processing process it is mandatory to manage it at any cost [[3]].

Defining the type, nature and volume of data at entry level by combining the all the aspect of healthcare in terms of data is very difficult task, this term can be referred as “Big Data”. Yet this is not a problem for most of the healthcare providers to manage data, but to achieve the QoS (Quality of Services) it is better to work for the future to handle data in new and improved way, that will surely help the medicos as well as patient in various ways. As per the market survey held by McKinsey, 71 to 100 billion could be saved by

implementing new ways of data analysis and management [[1]].

Hadoop or other related model is very much helpful to improve the outcome quality of data, Hadoop processing may help to stimulate computational methods available currently, and medical research approach to increase the output based quality [[6]]. To improve the quality of data only billing based data automation is not helpful. The emerging data based healthcare systems are Laboratory Information System, patients Personal Health Records. To address all these issue first we have to explore Data based QoS (Quality of Service) factors then security factors related to big data Analytics specifically for healthcare industry.

2. DATA BASED QOS (QUALITY OF SERVICE) FACTORS

In respect of health care data quality is must. Data quality means ‘How much data is fit for the purpose’. In order to offer a basic principle for analysis, a universally accepted set of basic data quality factors are required.

The upcoming section includes the sources available in the information system related literatures. Wang & Stornig (1996), had suggested two types of data quality factors first 'Intrinsic' and second 'Contextual'. ISO 8000 [[1]] series has also suggested various factors related to data.

A. Accuracy of Data

It can be defined as the degree of association of the data with the real-world atmosphere that it is intended to represent, normally quantified by typically a confidence interval, such as 'while measuring temperature of human body or any other species must be as accurate as within 1 degree Celsius.

B. Data Precision

During receiving data from external source, level of details must be measured such as 'whole numbers of degrees-Celsius'. This will be helpful to define the specific and valid domain of content for that data-item are defined, e.g. a numeric field may enable four digits after the decimal point, or integral value only, or multiples of 10 only; and YK2 Problem for century year change for the first time after computer machine was invented or birth date may identify DD-MM-YY (day-month-year), or MM-DD-MM-YY (Month-day-year), or year only [[5]].

C. Timeliness of Data

Timeliness of data can be defined with following elements:

- 1) *Temporal-Applicability*: This reflects, the set of data should be time stamped and must be discarded when time gets over, for example, the period during which an income-figure was earned, and the date after which a qualification or licence was applicable.
- 2) *Up-to-Datedness*: This reflects, the set of data should be updated time to time and the absence of up to date material lag between a real-world occurrence and responding data set.
- 3) *Currency*: This reflects when the data-item was recorded and was authenticated last time, or the period over which an average was computed. This also reflects volatile data-items, such as total rainfall for the last one month, total age, status, and fitness for work [[2]].

D. Data completeness

The availability of sufficient contextual information that the data is not liable to be misinterpreted [[6]].

3. DATA SECURITY FACTORS

The term data security is an essential for big data analytics

A. Data Consistency

While doing some same operation the users must receive the same set information as output, regardless the location from where that request was processed. It means that no matter which location that forms our data centre receives an order, everyone must respond to the operation transparently and equally. All clients must receive the same version of data [11].

B. Data Availability

The responding system must answer for the entire input request regardless of some responding nodes are not working properly.

C. Tolerance to Divisions

The system must be fault tolerant till maximum extent, even maximum part of the network is not working in order.

D. Data Integrity

Data integrity refers to the maintaining and assuring the accuracy and consistency of, data over its entire life-cycle.

4. ADVANTAGES OF BIG-DATA IN HEALTHCARE INDUSTRY

Big data analytics can help the healthcare industry in many ways:

- Big data could reduce the redundancy of data, and help the industry to avoid incorrect reporting, also help the industry to improve the response [[8]] [[9]].
- The Big data models of healthcare system can also incorporate real time information. Real time big data is advantageous in much respect. For example, any errors can easily be identified and trouble shoot in an organization immediately and the operational problem can be resolved [20].
- Big data will provide cost-effective and less time consuming system that will increase the productivity. The services also can be further improved with real time information assistance. For instance, it will provide the all patients related information and at the same time medical aid can be provided without any delay [[7]].
- In healthcare, big data based models are also used to address the incurable medical issues by predictive analysis. Healthcare professionals are able to reduce the risk and overcome the issue with the information gathered from the big data.
- Apart from that, big data is also helpful to identify frauds related to insurance claims etc. Fake and false claims can be identifiable. This

will facilitate insurance companies to prevent losses [[8]] [[9]].

- The big data will also helpful for development or invention for the target health problem [[20]].
- Clinical decision support system will also be assisted with the help of big data models. Big data will give closer look on population type and their related medical problem [[10]] [[11]].
- Big data could assist the pharmacy companies to discover new potential and effective drugs more quickly and deliver it to the patient very conveniently [[12]].

5. BIG DATA ISSUES FOR HEALTHCARE INDUSTRY

There are also a lot of challenges in big data while collecting, protecting and sharing health related data [[15]].

- Issues such as data privacy for patient and medicos, security related issues, standards governance related standards to be addressed [[16]].
- Healthcare data is not static, and most elements will require relatively frequent updates in order to remain current and relevant. For some datasets, like patient vital signs, these updates may occur every few seconds. [[17]] [[18]].
- The adverse effects of drugs use could also be determined, because all data are readily available everywhere [[19]].

6. CONCLUSION

Big Data is a growing field in healthcare outlook such as in drug discovery, patients personalization care, treatment efficiency, improvement in clinical outcomes, and patients safety management. Near in future we can also prepare new data and security models for measuring the security and quality of data using health care matrices. These models will be helpful to address new and challenging issues of big data in health care environment.

REFERENCES

[1] Better explaining the CAP Theorem – DZone Big Data". dzone.com. Retrieved 2016-09-02.

[2] IBM, Big data at the speed of business: Big data and analytics, Retrieved from URL <http://www-01.ibm.com/software/data/bigdata/>, September 2015.

[3] Harvard Business, How Big Data Impacts Healthcare by Harvard Business Review Analytics Services, 2014.

[4] Brian Eastwood, "Big Data Analytics Use Cases for Healthcare IT, [Personnel Survey] Unpublished raw data. Retrieved from <http://www.cio.com/>, 2013.

[5] Hongyong Yu, and Deshuai Wang, "Research and Implementation of Massive Health Care Data Management and Analysis Based on Hadoop", Fourth International Conference on Computational and Information Sciences, 2012.

[6] Lars George, Getting Started with Big Data Architecture. Cloudera, 2014. Retrieved from URL <http://blog.cloudera.com/blog/2014/09/getting-started-with-big-data-architecture/>

[7] Basco JA, Senthilkumar NC (2017) Real-time analysis of healthcare using big data analytics. IOP Conference Series: Materials Science and Engineering 263: 042056.

[8] https://www935.ibm.com/services/multimedia/Exploter_le_Big_Data_pour_lutter_contre_la_fraude_aux_sinistres_Juin_2013.pdf

[9] <https://www.insurancenixus.com/fraud/role-data-and-analyticsinsurance-fraud-detection>

[10] <https://www.scribd.com/document/273953518/Data-drivenHealthcare-Organizations-Use-Big-Data-Analytics-for-Big-Gains>

[11] <https://healthitanalytics.com/news/identifying-big-data-sourcesfor-population-health-management>

[12] <https://www.mckinsey.com/industries/pharmaceuticals-and-medical-products/our-insights/how-big-data-can-revolutionize-pharmaceutical-r-and-d>

[13] Kerremans GR (2016) Big data in Healthcare. J Health Commun 1: 4.

[14] Raghupathi W, Raghupathi V (2014) Big data analytics in healthcare: promise and potential. Health Inf Sci Syst 2: 3.

[15] Kerremans GR (2016) Big data in Healthcare. J Health Commun 1: 4.

[16] Raghupathi W, Raghupathi V (2014) Big data analytics in healthcare: promise and potential. Health Inf Sci Syst 2: 3. 2018 Vol.3 No.2:21-4 This article is available in: <http://healthcare-communications.imedpub.com/archive.php> ARCHIVOS DE MEDICIN

[17] Wakaskar RR, (2017), Brief Overview of Nanoparticulate Therapy in Cancer. J Drug Target 26: 123-126.

[18] Wakaskar RR (2017) Cancer Therapy with Drug Delivery Systems. J Pharmacogenomics Pharmacoproteomics 8: e158.

[19] Wakaskar RR, (2017), Challenges Pertaining to Adverse Effects of Drugs. Int J Drug Dev & Res 9: 01-02.

[20] Alexander CA, Wang L, (2017), Big data in healthcare: a new frontier in personalized medicine. Open Access J Trans Med Res 1: 0005.