RESEARCH ARTICLE OPEN ACCESS

Application of Big Data Analytics in Health Care

Mrs. Marrynal .S. Eastaff[1], Mrs. Premalatha .P[2]
[1]*PG Department of Information Technology, Hindusthan College of Arts and Science, Coimbatore, India*
[2]*Department of Information Technology and Computer Technology, Hindusthan College of Arts and Science, Coimbatore, India*

ABSTRACT
Big data literally means large collections of datasets containing abundant information. Big Data is now rapidly expanding in all science and engineering domains. The current trend in Big Data analytics and in particular Health Information Technology is towards building sophisticated models, methods and tools for business, operational and clinical intelligence. The healthcare industry has generated large amount of data generated from record keeping, compliance and patient related data. In today's digital world, it is mandatory that these data should be digitized. To improve the quality of healthcare by minimizing the costs, it's necessary that large volume of data generated should be analyzed effectively to answer new challenges. It requires a technology that helps to perform a real time analysis on the enormous data set. This paper provides an overview of big data analytics in healthcare systems. It describes about big data generated by these systems, data characteristics and how big data analytics helps to gain a meaningful insight on these data set.
Keywords: Big Data analytics, Data acquisition, Genomic Data, Predictive Data mining

I. Introduction

Big data analytics (BDA) has emerged from two distinct concepts – big data and analytics. Together it represents a new information management approach that has been designed to derive previously untapped intelligence and insights from data to address many new and important questions. Within the health sector, it provides stakeholders with new insights that have the potential to advance personalized care improve patient outcomes and avoid unnecessary costs. Analytics when applied in the context of big data is the process of examining large amounts of data, from a variety of data sources and in different formats, to deliver insights that can enable decisions in real or near real time. Big data analytical approaches can be employed to recognize inherent patterns, correlations and anomalies which can be discovered as a result of integrating vast amounts of data from different data sets.BDA can be used to help researchers find causes of, and treatments for diseases; actively monitor patients so clinicians are alerted to the potential for an adverse event before it occurs; and personalize care so precious resources associated with a treatment are not administered to a patient who cannot benefit from the intervention.

II. BDA Characteristics

Today, across all industries, the typical sources [2] of BD include:

- **Internet transactions** – in the near future more people will be online every day, including countless automated transactions creating a number of data points collected by retailers, banks, credit card issuers, credit agencies, social networking and search engine service providers and others.
- **Mobile devices** – There are more than 5.6 billion mobile phones in use worldwide. Each call, text and instant message is generating data. Mobile devices, particularly smart phones and tablets, also make it easier to use social networking and other data-generating applications.
- **Social networking and media –** There are currently millions active on Facebook,Twitter, and public blogs. Each Facebook update, tweet, blog post and comment creates multiple new data points – structured, semi-structured and unstructured – sometimes referred to as data exhaust.
- **Networked devices and sensors –** Electronic devices of all sorts – including servers and other IT hardware, smart energy meters and temperature sensors, patient monitors and aides – all create semi-structured log data that record every action.

III. Phases In The BDA Process

We can map steps taken up while performing BDA Process to the data mining knowledge discovery steps as follows:

1. Data acquisition and storage:

Data is fed to the system through many external sources like clinical data from Clinical Decision Support systems (CDSS), EMR, EHR, machine generated sensor data, data from wearable devices, drug related data from Pharmaceutical companies, social media data like twitter feeds, Facebook status, web pages, blogs, articles and many more.[3] This data is either stored in databases or data warehouse. With advent of cloud computing, it is convenient to store such voluminous data on the cloud rather than on physical disks. This is more cost effective and manageable way to store data.

2. Data cleaning:

The data which has been acquired should be complete and should be in a structured format, for performing effective analysis. Generally it is seen in that healthcare data has flaws like, many patients don't share their data completely like data about their dietary habits, weight and lifestyle. In such cases the empty fields need to be handled appropriately. The data from sensors, prescriptions, medical image data and social media data need to be expressed in a structured form suitable for analysis.[1]

3. Data integration:

The BDA process uses data accumulated across various platforms. This data can vary in metadata (the number of fields, type, and format). The entire data has to be aggregated correctly and consistently into a dataset which can be effectively used for data analysis purpose [8].

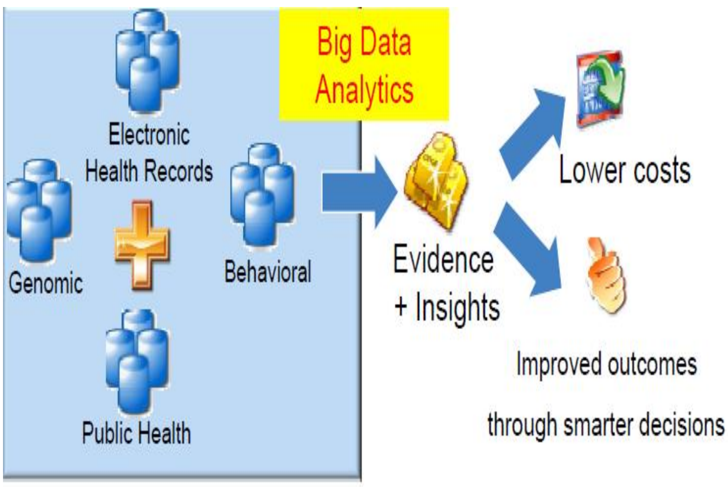
4. Data querying, analysis and interpretation:

Once the data is cleaned and integrated, the next step is to query the data. A query can be simple query or a complex one. Depending on the complexity of the query, the data analyst has to choose appropriate platform and analyze. Some of the platform and tools are Hadoop, MapReduce, Storm, GridGrain. Big data databases like assanadra, HBase, MongoDB, CouchDB, OrientDB, Terrastore, Hive etc. Data Mining tools like RapidMiner,Mahout, Orange, Weka, Rattle, KEEL etc. Programming languages like Pig/PigLatin, R, ECL. Big data search tools like Lucene, Solr etc.Data Aggregation and transfer tools like Sqoop, Flume,Chukwa[3] .

IV. Opportunities For BDA In Healthcare

Big data analytics represents a new approach to analytics. Some of the innovative ideas and solutions are:

- Clinical decision support – BDA technologies that sift through large amounts of data, understand, categorize and learn from it, and then predict outcomes or recommend alternative treatments to clinicians and patients at the point of care.
- Personalized care – Predictive data mining or analytic solutions that can leverage personalized care (e.g., genomic DNA sequence for cancer care) in real time to highlight best practice treatments to patients. These solutions may offer early detection and diagnosis before a patient develops disease symptoms.
- Public and population health – BDA solutions that can mine web-based and social media data to predict flu outbreaks based on consumers' search. BDA solutions can also support clinicians and epidemiologists to help identify disease trends [4].
- Clinical operations – BDA can support initiatives such as wait-time management, where it can mine large amounts of historical and unstructured data, look for patterns and model various scenarios to predict events that may affect wait times before they actually happen.
- Policy, financial and administrative – BDA can support decision makers by integrating and analyzing data related to key performance indicators.



V. Leveraging Big Data In Healthcare

By discovering associations and understanding patterns and trends within the data, big data save lives. When big data is synthesized and analyzed, healthcare providers, drug manufacturers, and other stakeholders in the healthcare industries can develop more thorough and insightful diagnoses and treatments, delivering higher quality care at lower costs and achieve better outcomes overall. The potential for big data analytics in healthcare lead to better outcomes across many scenarios.

Healthcare and Payers

- Analyzing patient characteristics and the cost to identify the most clinically effective and cost-efficient diagnoses and treatments.
- Identifying, predicting, and minimizing fraud by implementing advanced analytic systems for fraud detection and checking the accuracy and consistency of claims.
- Analyzing large numbers of claim requests rapidly in the pre-adjudication phase to reduce fraud, waste and abuse [5].

Evidence-Based Medicine

- Combining and analyzing a variety of structured and unstructured data – EMRs, financial and operational data, clinical data, and genomic data – to match treatments with outcomes, predict patients at risk for disease or readmission, and provide more efficient care at reduced cost.
- Applying advanced analytics to patient profiles (e.g., segmentation and predictive modeling) to identify individuals who would benefit from proactive care or lifestyle changes
- Using historical data to personalize medical care by predicting and/or estimating developments or outcomes, such as which patients will choose elective surgery, will not benefit from surgery, are at risk for medical complications or hospital-acquired illness, or will have possible co-morbid conditions.
- Executing gene sequencing more efficiently and cost effectively to make genomic analysis a part of the regular medical care decision process and the growing patient medical record.

Real-Time Healthcare and Clinical Analytics

- Aggregating and synthesizing patient clinical records and claims datasets in real time to provide data and services to third parties.
- Detecting individual and population trends more rapidly and accurately by developing and deploying mobile applications that help patients manage their care, locate providers, and improve their health.
- Monitoring medical devices, including wearable’s, to capture and analyze in real-time large volumes of fast-moving data, for safety monitoring and adverse event prediction, enabling payers to monitor adherence to drug and treatment regimens and detect trends that lead to individual and population wellness benefits [5].

Public Health

- Analyzing disease patterns and tracking disease outbreaks and transmission to improve public health surveillance and speed response.
- Improving data models to better predict virus evolution, leading to more accurately targeted seasonal vaccines (e.g., choosing the annual influenza strains).
- Turning large amounts of data into actionable information that can be used to identify needs, provide services, and predict and prevent crises.

VI. Big Data Challenges In Health Care

- Inferring knowledge from complex heterogeneous patient sources.
- Leveraging the patient/data correlations in longitudinal records.
- Understanding unstructured clinical notes in the right context.
- Efficiently handling large volumes of medical imaging data and extracting potentially useful information and biomarkers [6].
- Analyzing genomic data is a computationally intensive task and combining with standard clinical data adds additional layers of complexity.
- Capturing the patient’s behavioral data through several sensors; their various social interactions and communications [7].

VII. Conclusion

Big data analytics has the potential to transform the way life sciences and healthcare organizations use sophisticated technologies to gain insights from their clinical and other data repositories to make informed decisions. Analytics allow organizations to investigate and explore data to identify relationships, trends, and patterns to reveal insights that, when combined with business context, create knowledge. In the future, the implementation and use of big data analytics will spread rapidly.

References

[1] *Content and Predictive analytics in Healthcare, IBM - February 2012.*

[2] *Estimates come from a variety of research, including Gartner, IDC, Forrester, TDWI and materials from IT vendors such as IBM, ORACLE, SAP, SAP, Cloudera, Greenplum and Vertica.*

[3] *Cynthia Harvey, "50 Top Open Source Tools for Big Data ", Available: http://www.datamation.com/data-center/50-top-open-source-tools-for-big-data-1.html*

[4] *"Big Data Analytic in Health" White paper (Full Report, April 2013.*

[5] *https://knowledgent.com/whitepaper/big-data-analytics-life-sciences-healthcare-overview/*

[6] *A Survey On Big Data Analytics In Health Care, P r i y a nka K et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (4) , 2014, 5865-5868 https://www. siam. org/meetings/sdm13/sun.pdf*

[7] *Volume 6, Issue 4,April 2016 ISSN: 2277 128X International Journal of Advanced Research in Computer Science and Software Engineering "Big Data Analytics in Healthcare-Its Benefits, Phases and Challenges" Jasleen Kaur Bains*

[8] *Wullianallur Raghupathi and Viju Raghupathi, "Big data analytics in healthcare: promise and potential", Health Information Science and Systems 2014, 2:3,Available: http://www. hissjournal. com/content/2/1/3*

[9] *D. Agrawal et. al, "Challenges and Opportunities with Big Data " , Big Data White Paper - Computing Research Association, Feb 2012. Available: http://cra. org/ccc/ docs/init/ bigdatawhitepaper. pd*

[10] *https://knowledgent.com/whitepaper/big-data-analytics-life-sciences-healthcare-overview/ H.Hu,Y. Wen,T. -S .Chua,andX.Li, ''Towards scalable systems for big data analytics: A technology tutorial,'' IEEE Access , vol. 2, pp. 652–687, 2014.*

[11] *S. LaValle et al., ''Big data, analytics and the path from insights to value,'' MIT Sloan Manage. Rev.,vol.21,2013.*

[12] *H. Moed et al., ''Special issue on big data,'' Res. Trends Sep.30,2012.*

[13] *IBM Enterprise Content Management – ECM(2014), available at: http://www-03.ibm. com/software/products/en/category/enterprise-content-management (accessed 12 July 2014).*

[14] *Huberman B. 2012, Sociology of science: Big data deserve a bigger audience, Nature, vol. 482, pp.308.*

[15] *Luo et al. 2012, Dijun Luo, Chris Ding, Heng Huang, Parallelization with Multiplicative Algorithms for Big Data Mining, In: Proc. of IEEE 12th International Conference on Data Mining, pp.489-498, 2012*