ISSN:0975-9646

P r i y a nka K et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (4) , 2014, 5865-5868

A Survey On Big Data Analytics In Health Care

Priyanka K
B.V.B C.E.T
Hubli,

Prof Nagarathna Kulennavar
B.V.B.C.ET
Hubli,

Abstract- **This paper gives a brief introduction about how we can uncover additional value from health information used in health care centers using a new information management approach called as big data analytics .Including big data analytics in health sector provides stakeholders with new insights that have the potential to advance personalized care, improve patient outcomes and avoid unnecessary costs. This paper defines big data analytics and its characteristics, comments on its advantages and challenges in health care.**

I. Introduction

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 toward rapid digitization of these large amounts of data. Driven by mandatory requirements and the potential to improve the quality of healthcare delivery meanwhile reducing the costs, these massive quantities of data (known as 'big data') hold the promise of supporting a wide range of medical and healthcare functions, 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.

By definition, big data in healthcare refers to electronic health data sets so large and complex that they are difficult (or impossible) to manage with traditional software and/or hardware; nor can they be easily managed with traditional or common data management tools and methods. Big data in healthcare is overwhelming not only because of its volume but also because of the diversity of data types and the speed at which it must be managed.

It includes clinical data and clinical decision support systems (physician's written notes and prescriptions, medical imaging, laboratory, pharmacy, insurance, and other administrative data); patient data in electronic patient records (EPRs); machine generated/sensor data, such as from monitoring vital signs; social media posts, including Twitter feeds (so-called tweets), blogs , status updates on Facebook and other platforms, and web pages; and less patient-specific information, including emergency care data, news feeds, and articles in medical journal.

For the big data scientist, there is, amongst this vast amount and array of data, opportunity. By discovering associations and understanding patterns and trends within the data, big data analytics has the potential to improve care, save lives and lower costs. Thus, big data analytics applications in healthcare take advantage of the explosion in data to extract insights for making better informed decisions.

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. Various analytical concepts such as data mining, natural language processing, artificial intelligence and predictive analytics can be employed to analyze, contextualize and visualize the data. 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.

This paper provides an overview of big data analytics in healthcare as it is emerging as a discipline. First, we define and discuss the definition of big data and characteristics of big data analytics in healthcare. Then we describe the types of big data in healthcare. Third, we provide examples of big data analytics in healthcare. Fourth, the challenges are identified. Lastly, we offer conclusions and future directions.

II. Defining Big Data.

Big data typically refers to the following types of data:

- Traditional enterprise data – includes customer information from CRM systems, transactional ERP data, web store transactions, and general ledger data.
- Machine-generated /sensor data – includes Call Detail Records ("CDR"), weblogs, smart meters, manufacturing sensors, equipment logs (often referred to as digital exhaust), and trading systems data.
- Social data – includes customer feedback streams, micro-blogging sites like Twitter, social media platforms like Facebook

In fact, there are four key characteristics that define big data:

- **Volume** is the amount of data generated by organizations or individuals. Enterprises in all industries are looking for ways to handle the ever-increasing data volume that's being created every day.
- **Velocity** is the frequency and speed at which data is generated, captured and shared. Consumers as well as businesses now generate more data and in much shorter cycles, from hours, minutes, seconds down to milliseconds.
- **Variety** is the proliferation of new data types including those from social, machine and mobile sources. New types include content, location or geo-spatial, hardware data points, log data, machine data, metrics, mobile, physical data

points, process, radio frequency identification (RFID), search, sentiment, streaming data, social, text and web. Also, variety includes traditional unstructured clinical data (i.e., free text).

III. BIG DATA IN HEALTH CARE

The types of data anticipated to be of use in BDA include:

1. Clinical data – up to 80 per cent of health data is unstructured as documents, images, clinical or prescribed notes;
2. Publications – clinical research and medical reference material;
3. Clinical references – text-based practice guidelines and health product (e.g., drug information) data;
4. Genomic data – represents significant amounts of new gene sequencing data;
5. Streamed data – home monitoring, telehealth, handheld and sensor-based wireless or smart devices are new data sources and types;
6. Web and social networking data – consumer use of Internet – data from search engines and social networking sites; and
7. Business, organizational and external data – administrative data such as billing and scheduling and other non-health data.

IV. OPPORTUNITIES OF BDA IN HEALTH CARE

Big data analytics represents a new approach to analytics. It does not yet have a large or significant footprint India or internationally. However, the continuing digitization of health records together with the interoperable electronic health record (EHR), presents new opportunities to investigate a myriad of clinical and administrative questions.

There is potential to layer BDA-type applications, in a privacy-protective manner, on top of the foundational health IT infrastructure to derive value that might not otherwise be found. What follows are some innovative ideas and solutions.

- 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, social content and query activity. BDA solutions can also support clinicians and epidemiologists performing analyses across patient populations and care venues to help identify disease trends.
- 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. IMPLEMENTATION PROGRESS OF BDA WITHIN THE HEALTH CARE SECTORS.

Example1: Disease outbreak

BDA used for monitoring of disease networking. An example is Google. Org's use of BDA to study the timing and location of search engine queries to predict disease outbreaks. Research shows that one-third of consumers currently use social networking for health care purposes (Facebook, YouTube, blogs, Google, Twitter). As demands for access to health information from social networking sites continue to proliferate, BDA can potentially support key prevention programs such as disease surveillance and outbreak management.

Example 2: Question and answer – clinical decision support

Research indicates that 79 per cent of provider organizations in the U.S. are turning to clinical informatics in an effort to prevent medical errors. 61 per cent expect analytics to improve population health, and 52 per cent indicate that analytics-driven preventive care will help rein in costs. BDA can potentially help improve existing workflow and outcomes from business processes such as appointment brokering, scheduling, e-referral and e-discharge.

BDA can also assist in providing insights around gaps in the continuum of care across settings and highlight best practices in care processes and clinical outcomes.

Question and answer solutions integrated with computerized provider order entry (CPOE), e-referral, e-discharge and other process based activities can be used to analyze and predict trends in health care.

BDA can mine volumes of medical literature and other unstructured data and integrate these results with the increasing volumes of discrete data captured in EHRs, EMRs and PHRs. BDA can combine content analysis, evidence-based data and through natural language processing technology can understand, learn and then predict future events. These analytics are then fed back to clinicians as considerations in their decision making.

Patients or consumers also use BDA to get answers for their own conditions. Data could be presented back in a meaningful way and encourage patient participation in their health care plans and potentially reduce re-admissions or adverse outcomes.

Example 3: HERITAGE HEALTH PRICE

Identify patients who will be admitted to a hospital within the next year using historical claims data.

Over $30 billion was spent on unnecessary hospital admissions.

Goals:

- Identify patients at high-risk and ensure they get the treatment they need.
- Develop algorithms to predict the number of days a patient will spend in a hospital in the next year.

Outcomes:

- Health care providers can develop new strategies to care for patients before its too late reduces the number of unnecessary hospitalizations.
- improving the health of patients while decreasing the costs of care.
- Winning solutions use a combination of several predictive models.

Example 4: DATA STREAMING

BDA solutions are providing early or predictive insights for clinicians and patients about treatment compliance and adverse events. The example shows the use of BDA at Toronto's Hospital for Sick Children to synthesize the deluge of information that monitors capture from neonates (more than 1,000 recordings per second of physiological measures such as body temperature, heart rate, respiratory rate and blood pressure). The BDA solution provides insights which allow researchers to create algorithms to predict when a baby is at risk of infection. As personal health devices and mobile sensor applications continue to proliferate, more information will become available as a result that can employ BDA.

Example 5: GENOMICS AND PERSONALIZED CARE

Personalized care is being presented as the next wave of transformation in the delivery of medical treatment to patients.

The world's largest set of data on human genetic variation – produced by the international 1000 Genomes Project – is now freely available on the Amazon Web Services (AWS) cloud. At 200 terabytes – the equivalent of 16 million file cabinets filled with text, or more than 30,000 standard DVDs – the current 1000 Genomes Project dataset is a prime example of BD, where datasets become so massive that few researchers have the computing power to make the best use of them. AWS is storing the 1000 Genomes Project as a publicly available dataset for free, and researchers will pay only for the computing services they use.

Example 6: PENALTIES FOR POOR CARE -30 DAYS READMISSIONS

Hospitalizations account for more than 30% of the 2 Trillion annual cost of healthcare in the United States. Around 20% of all hospital admissions occur within 30 days of a previous discharge. Not only expensive but are also potentially harmful, and most importantly, they are often preventable.

- Medicare penalizes hospitals that have high rates of readmissions among
Patients with heart failure, heart attack, and pneumonia.
- Identifying patients at risk of readmission can guide efficient resource utilization and can potentially save millions of healthcare dollars each year.
- effectively making predictions from such complex hospitalization data will require the development of novel advanced analytical models.

Example 7: Consumer based social media

This example involves use of social media and networks to inform consumers of conditions and self-management recommendations. The example depicted to the left is the integration and provisioning of population based information within social networking environments to better inform and engage patients in managing their health and to influence or
change their habits for conformance to treatment options, plans and best practices.

However, the use of social network search engine toolkits does pose some interesting legal or liability considerations. Vendors of these toolkits may not provide evidence-based recommendations and treatment options to patients. In fact, patients may be at risk when they self-diagnose based on these recommendations.

Example 8: Supporting health innovation through the use of open health data

The U.S. Open Health Data Initiative is a public-private effort that aims to help Americans understand health and health care performance in their communities. This work centers on catalyzing the advent of a network of community health data suppliers (HHS) and "data appliers" who utilize that data to create BDA applications that raise awareness of community health performance, increase pressure on decision makers to improve performance, and help facilitate and inform action to improve performance.

Working with a growing array of health care stakeholders, they will be seeking to identify the uses of this data that would do the most to raise awareness of health performance, and help motivate civic leaders and citizens to improve performance.

Potential examples of uses include:

- Interactive health maps on the web that allow citizens to understand health performance in their area versus other areas, with tremendous ease and clarity;
- "dashboards" that enable mayors and other civic leaders to track and publicize local health performance and issues;
- Social networking applications that allow health improvement leaders to connect with each other, compare performance, share best practices, and challenge each other;
- Competitions regarding how communities can innovate to improve health performance;
- viral online games that help educate people about community health;
- utilization of community health data to help improve the usefulness of results delivered by web search engines when people do health-related searches and further raise awareness of community health performance; and
- integration of community health-related data into new venues, such as real estate websites, which could be highly effective disseminators of such information.

VI. Big Data Challenges In Health Care.

- 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.
- 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.

VII.Conclusion And Future Work

Big data analytics in healthcare is evolving into a promising field for providing insight from very large data sets and improving outcomes while reducing costs. Its potential is great; however there remain challenges to overcome.

Big data analytics has the potential to transform the way healthcare providers use sophisticated technologies to gain insight from their clinical and other data repositories and make informed decisions. In the future we'll see the rapid, widespread implementation and use of big data analytics across the healthcare organization and the healthcare industry. To that end, the several challenges must be addressed. As big data analytics becomes more mainstream, issues such as guaranteeing privacy, safeguarding security, establishing standards and governance, and continually improving the tools and technologies will garner attention. Big data analytics and applications in healthcare are at a nascent stage of development, but rapid advances in platforms and tools can accelerate their maturing process.

Acknowledgment

The author would like to thank Prof Nagarathna Kulennavar for support used to conduct this work.

References

1. *Dembosky A:* "Data Prescription for Better Healthcare." Financial Times, December 12, 2012,
2. *Feldman B, Martin EM, Skotnes T:*"Big Data in Healthcare Hype and Hope." October 2012. Dr. Bonnie.
3. *Fernandes L, O'Connor M, Weaver V: Big data, bigger outcomes.*J AHIMA *2012, 38-42.*
4. Transforming Health Care through Big Data Strategies for leveraging big data in the health care industry. *2013.*
5. *Wullianallur Raghupathi and Viju Raghupathi Big data analytics in healthcare: promise and potential*