

A Survey: Big Data Analytics on Healthcare System

R. Sathiyavathi

Faculty of Computing, Sathyabama University, Chennai- 600119, India

Copyright © 2014 R. Sathiyavathi. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Big data and the related technologies have improved health care enormously, from understanding the origins of diseases, better diagnoses, helping patients to monitor their own conditions. By digitizing, combining effectively using big data, healthcare organizations can improve their quality of service by analyzing the effectiveness of a treatment and also the efficiency of the healthcare delivery process by detecting fraud, waste and drug abuse more quickly and efficiently. General goals to use analytics are, we can predict readmission risks, increase the efficiency of clinical care, and finding opportunities for cost savings. This paper gives various solutions for how and where big data can be applied in the health care system.

Keywords: Big Data, HealthCare Solutions, Hadoop Use Cases, Clinical Decision Support

1. Introduction

Big data is a collection of techniques and technologies which needs new forms of integration to uncover large hidden values from large datasets that are diverse, complex, and of a massive scale. Big data can also be defined as large volume of unstructured data which cannot be handled by traditional data management tools like relational database management system. In 2012, Gartner s defined big data as follows: Big data are huge volume, high velocity, and large variety information assets that require new forms of processing to enable strong decision making, insight discovery and process optimization. Additionally, another V is added called veracity, shows the uncertainty of the data, as it usually comes from a variety of sources like as given below [3].

- Web and social media data - interaction data like Facebook, Twitter and other blogs etc.
- Electronic Health Records (EHR) - Health claims and other records that can be structured or unstructured.
- Reading from sensor, meters and other devices.
- Biometrics - Fingerprints, genetics, handwriting and retinal scan data, it consists of X-rays and other medical images and blood pressures and pulse and similar type of data.
- Human generated data - electronic mail records (EMR), physician notes, paper documents etc.

The increasing digitization of healthcare information is opening new possibilities for providers and payers to enhance the excellence of care, improve healthcare outcomes, and reduce costs. Due to advance technologies the paper works are converted into digital format (digital health records or Electronic Health Records (EHR). Since information is in digital form, healthcare providers can use some available tools and technologies to analyze that information and generate valuable insights.

360-degree view for every patient is created by integrating electronic health records (EHRs), videos, medical images, scanned documents, and physicians' notes. By integrating various social media details, demographic details, environmental information, and behavioral information related to patients, organizations are able to discover new correlations. Creating a more holistic view of each patient and analyzing a wider array of information will help organizations meet the requirements of emerging healthcare models.

2. Literature Survey

The increasing digitization of healthcare information is opening new possibilities for providers and payers to improve the quality of care, health care results, and minimize the costs. The latest tools and technologies are used on digital information of health care organizations can generate valuable insights. Organizations must also analyze internal and external patient information to more accurately measure risk and outcomes. At the same time, many providers and payers are working to increase data transparency to produce new insight knowledge.

Integrated delivery networks can be formed using health information exchange by various providers and payers. Some large pharmaceutical companies are de-identifying data from their clinical trials, protecting patients' privacy while making data available to qualified researchers outside the organization. Kiyana Zolfaghar et al. [1] has presented prediction model for the risk of readmission of

congestive heart failure incidents, where Mahout framework is used. Training data is given to the Hadoop file system, raw data is preprocessed and converted to classifiable data which will be in encoded format called as a vector that will be given as input to the Mahout framework, using random forest algorithm.

Joseph M. Woodside [2] has presented, inefficient vendors can be identified, who is poor in the member's lifestyle decisions and compliance with preventative care programs. For individuals, intensive can be given, such as cash, gift cards which is considered as one of big recommended change in the health care system.

Existing analytical techniques can be applied to the vast amount of existing (but currently unanalyzed) patient related health and medical data to reach a deeper understanding of results, which can be applied at the point of care. Ideally, these data would inform each physician and their patients during the decision-making process and used to Identify the appropriate treatment option for that particular patient

3. Tools and Application in Health Care System

The health care system has a large volume of unstructured data, so it is impossible to do research and diagnoses without an appropriate tool or technique. Hadoop is a tool that is designed to process huge volumes of data, which is integrated with map-reduce concept. Map reduce can divide the data set into multiple chunks, each will be processed in parallel among multiple nodes. MapR can overcome the limitation of Hadoop, as it has dynamic read-write data layer that provides unparalleled dependability. MapR contains the following packages which provides unique strength to the system - Hbase, Hcatalog, Oozie, Mahout, Sqoop , ApacheHive, Apache Pig, Cascading etc., in MapR platform analytics can be done directly on the data.

A. Application Of Big Data In Health Care:

1. **Personalized Treatment Planning:** Based on the medical histories of every individual patient, diagnoses can be done, which can be used to decide the appropriate treatment and medicine for that patient. Real time analysis will be done using MapR and Hadoop, based on the analytics results, the patient can have personalized care for them.

2. **Utilization Review:** To assist evidence-based treatment, which is considered to be the best form of treatment, the big data analytics of health information are required. One such is the retrospective drug utilization review, which gives insightful information on how effective a particular drug is in the treatment of a certain health condition. The Pharmacy and Therapeutics committee (P&T committee) of every payer organization constantly review the utilization patterns of the patients by generating various analytical reports from their medical and drug claims. This analysis can be further improved by getting information from non-traditional sources like social and other electronic media for more insightful information using big data analytics tools and techniques like Hadoop and MapReduce.

3. **Assisted Diagnosis:** Physicians can isolate and treat the patient based on some factors like symptoms, medical history, and side effects. Using prediction modeling and machine learning, Hadoop can provide information which will be helpful to the doctors.
4. **Fraud Detection:** Fraud should be avoided in health care system to reduce the cost for both the insurance companies and the patients. The existing techniques using the traditional healthcare analytics are retrospective in nature in analyzing the fraud. With the help of big data analytics using certain tools and techniques, certain patterns can be detected on a more real-time basis and can be designed to intimate the payer organization immediately on any fraud or suspected activity, so that they can do the right intervention proactively.

B. Prediction Algorithm:

The first step is a collection of data from various data sources, prediction attribute will be identified, and for a particular case the set of prediction attributes will be collected, the respective algorithm should be applied in case the attribute. Then contextualization can be applied and desired performance can be retrieved. The prediction algorithm is given in the figure 1.

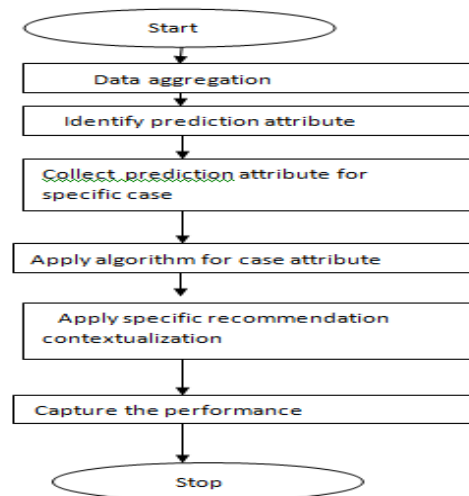


Fig. 1. Prediction Algorithm

4. Conclusion

The implementation of new healthcare regulations and the availability of financial incentives are leading Healthcare organizations to embrace a digital environment for healthcare information. With the digitization of health information, they can generate deep insights that can streamline clinical workflows, optimize care, strengthen doctor-patient relationships, cut costs, and improve outcome. Due to the

very nature of the Healthcare information being too huge and unstructured, it is pertinent to apply the big data analytics to generate such insightful information.

Acknowledgements. We thank Sathyabama University for providing us with various resources and an unconditional support for carrying out this work.

References

[1] Kiyana Zolfaghar, Naren Meadem, Ankur teredesai, Senjuti Basu Roy, Si-Chi Chin. Big Data Solutions for Predicting Risk-of-Readmission for Congestive Heart Failure Patients. 2013 IEEE International Conference on Big Data, 978-1-4799-1293-3/13. <http://dx.doi.org/10.1109/bigdata.2013.6691760>

[2] Joseph M. Woodside. Virtual Health Management, 2014 11th International Conference on Information Technology: New Generations 978-1-4799-3187-3/14. <http://dx.doi.org/10.1109/itng.2014.124>

[3] Wullianallur Raghupathi, Viju Raghupathi R. Big data analytics in healthcare: promise and potential. Health Information Science and Systems, 2014. <http://dx.doi.org/10.1186/2047-2501-2-3>

[4]<http://www.insight.com/insighton/healthcare/capitalizing-power-big-data-healthcare>

[5] <https://hbr.org/2014/12/why-health-care-may-finally-be-ready-for-big-data>

[6]<http://www.intel.com/content/www/us/en/healthcare-it/sas-big-data-digitization-of-healthcare-white-paper.html>

Received: December 27, 2014; Published: February 9, 2015