

Overview of Big data's Contribution in Health Care

Anusha S K¹, Dhruti H G² & Mrs. Smitha G R³
^{1,2,3}Dept. of ISE, R V College of Engineering, Bengaluru, India

Abstract— *Big data is used to describe large quantity of both structured and unstructured data. Big data finds its applications in various fields. Here, in this paper we have concentrated on the application of big data in health care sector. Hospitals contain large volumes of patient's data. Incorporating the concepts of big data in healthcare, the patients can be benefited in many ways.*

Keywords— *Big data, healthcare, electronic health record, benefits, tools and technologies.*

1. Introduction

Big data is a term used to describe a massive volume of data that is very large and complex. Therefore it is difficult to process using traditional database and software techniques.

Every day, lots of data is being generated in all dimensions of healthcare like patient records, prescriptions, lab test reports etc...The biggest sources of big data in healthcare are: Hospitals, Health insurance companies, Research institutions, Pharmaceuticals and Medical devices. In order to improve the quality of healthcare by minimizing the costs, it is necessary that the data generated should be analyzed effectively to answer new challenges and to improve the healthcare services like providing customized patient treatment, early prediction of diseases etc. In reference to the above, this paper aims to provide deeper insight of incorporating big data in health care sector.

2. Properties of Big Data

The 5 types of properties identified are as follows [1]:

2.1. Volume: The health care data grows rapidly

Huge amount of data are collected by healthcare systems. The information stored in the healthcare scenario could be: personal patient data, medication history, symptoms, X-ray images, ECG reports, radiology reports etc... These systems usually use cloud for storage and processing of voluminous data.

2.2. Variety: The data comes from various sources

The healthcare data is usually semi-structured or unstructured. This is challenging as the data is both voluminous and is not structured. Therefore, handling such data efficiently to get value out of it is a big challenge.

2.3. Velocity: The speed of data generation

The health care systems usually store correct data; however it may not always be accurate. Also, there is a need to update it on a regular basis. Accuracy is a very important factor here along with the updating time. The healthcare system should ensure that the data is updated as soon as it arrives into the storage medium. Life or death of patients can rely on real time analysis of the healthcare data. Therefore, big data analytics must be performed in real-time to ensure proper treatment to the patients.

2.4. Veracity: The healthcare data has different meaning in different contexts.

High-quality and accurate data is essential for effective decision making. In such a case, ensuring uniform quality across the whole database is not possible. The quality of data is also an important factor in analysis and decision making. Furthermore, the survival of a patient may depend on the decision made using this data. So, the healthcare systems should ensure some good quality of data.

2.5. Value: Outcome of analytics.

Value is the outcome of processing the big-data. In healthcare field, value can be interpreted as the analysis resulting in prediction of future health issues based upon the currently available data, finding the best possible medications, treatments and therapies for the patient, prescribing preventive medicines, enhancing patient's health based on the analysis, reducing cost etc...Furthermore, it should improve the quality of service delivered by the healthcare organization. This will in turn increase the trust of the healthcare service consumers of the organization.

3. Need for Big Data

3.1. Providing patient centric services

Organizations always aim towards providing better user satisfaction. Similarly in healthcare sector, the aim is to make things better for the patient. Big data analytics when used in healthcare could be beneficial to patients as they need not carry all their health reports every time they go for consultation as the data would be digitized. In addition to this, the big data healthcare systems makes the job of the doctor too easier as they need not go through all medical records, but can request only a summarized report of the patient's health records using the analysis tools. This saves time and effort of both the patient and the doctor [2].

3.2. Predicting diseases earlier:

If the health records of a person are maintained, then it would be easy to judge his current health state and predict any further medical complications that could be caused. So, the doctor would be well aware of the diseases that may affect the person and therefore he/she can take preventive actions by prescribing proper medication to the patient. This will reduce the number of visits to the hospital.

3.3 Providing customized treatment

The effect of medicine on man's body is analyzed. Based on the analysis, further dosage can be altered. For example, the blood pressure of a person is tracked for a certain period of time and dosage of medicine is calculated for further treatment. This also helps the doctors to treat other patients who show similar symptoms.

4. Applications in Health Sector

4.1. Analysis of data collected by specialized sensors

Wearable devices are being widely used in the medical field. The devices consist of specialized sensors that collect data by measuring various physical parameters of a patient like, heartbeat, pulse rate, temperature etc. The data collected by such sensors can be analyzed to predict certain diseases or certain biological conditions of the patient.

4.2. Sleep monitoring

Many new apps and devices are being used that help in monitoring the sleep patterns of a patient. They could collect data like, the total sleep duration of the patient, the total REM sleep time of the patient etc... This data when analyzed can help the doctors treat insomnia patients and consequently used to treat patients who have psychological problems.

4.3. Fitness Tracking

People are becoming more and more aware of the advantages of being fit, new apps and devices are now available to track the fitness routines of a person. This may keep track of the physical activity done by the person like the amount of time spent in jogging; the increase in heartbeat during the fitness regime, the amount of calories burnt etc... The analysis of this data can be very useful for doctors to treat lifestyle disorder issues of a patient.

4.4. Mood prediction

Depression is a common psychological issue in the world. The biological diseases can be easily detected by the symptoms, but psychological issues are hard to detect in its initial stages. So, mood sensors can be helpful in detecting such problems and analysis of the mood patterns of a patient suffering from psychological disorders can help psychologists and psychiatrists to provide them with timely counseling and medication.

4.5. Evidence based medicine:

Evidence-based medicine (EBM) is an approach that optimizes the decision-making process by making use of evidence from well designed and conducted research. The decisions and policies are based on evidence. Thus it assures that the knowledge gap of an expert medical practitioner's opinion is supplemented with all available knowledge from the available medical literature so that best practice can be found and applied. The decision making process can make use of big-data analysis.

4.6. Electronic health record

EHR are the records that contain the details of the patients in digital format. These details may be shared across multiple health care centers, physiotherapists and personal trainers who will be able to guide the patients based on the information present in the EHR [3][4].

4.7. Telemedicine

It is a technology which provides medical aids to a patient over long distances. It facilitates the doctor to provide medical assistance to the patients who are in different locations just by examining the EHR of the patient. Further, assistance can be provided by various communication means like video chat, messages, over the telephone, etc.

5. Benefits

5.1. Patient's perspective

Big-data in healthcare helps the patients in understanding their health state better. With the use of EHR, telemedicine and similar such technologies the health care sector is now much closer to the patient and it is becoming more friendly. Applying

big data analytics on human genome mapping helps in treating cancer.

5.2. Doctor's perspective

The use of big data has revolutionized the scope for improvement. With more detailed data available, the performance of doctors is greatly improved. Furthermore, with the help of ongoing analysis and reporting facilities, the doctors are able to evaluate themselves. The capability of the healthcare system to store and process huge medical data helps the doctor to understand the complete history of patients and the doctor is able to give precise medications to the patient.

5.3. Researcher's perspective

With the help of big data the researches are able to mine the data and use it to analyze the trends in the field of medicine. Using the legacy data, they can predict the diseases that may occur in the future. For example, the algorithms of data mining and machine designs are useful to detect influenza epidemics by applying queries. In this way, researchers are able to contribute more algorithms and tools and also early predictions helps in improving the services of health sector.

5.4. Pharmaceutical company's perspective

The use of big data has greatly helped the pharmaceutical companies to produce the right quantity of drugs at right time. For example, an asthma patient will suffer from asthma attack comparatively high in the season where the content of pollen grains is high in the environment. Consequently, during that season, the pharmaceutical companies manufacturers more asthma related drugs.

6. Technological Solutions for Healthcare Sector

Most of the data in healthcare sector is unstructured as the medical records majorly consists of personal case records of the patients, x-rays, ultrasound images, Doppler images, prescriptions, lab test report documents, ECG graphs and reports, MRI scan images, data collected by the wearable medical devices etc.. The data is huge, rapidly changing and needs real-time processing capabilities for decision making. Some available technologies and tools for the purpose of analysis and storage of big data are as listed below [5].

6.1. Hadoop

Apache Hadoop is an open-source framework for distributed storage and processing of big-data. The modules in Hadoop are designed with fault-tolerance capability.

The core consists of a storage which is known as Hadoop Distributed File System (HDFS), and a processing part is the Map Reduce. Hadoop splits the data in HDFS into big blocks and distributes them across nodes in the cluster. The processing of data is done in parallel on the nodes. The data locality is effectively utilized as the nodes usually process data that is accessible by them only. Thus, improving the processing speed.

The base framework consists of the following modules, but many other additional modules can be installed on top of it as requirements arise.

- *Hadoop Common* – It contains libraries and utilities that are needed by the other Hadoop modules.
- *Hadoop Distributed File System (HDFS)* – It is a distributed file-system that stores data on different machines in a cluster and provides a very high bandwidth for movement of data across the cluster.
- *Hadoop YARN* – It is a resource-management platform that is responsible for the management of computing resources in clusters so that, the resources are efficiently utilized to schedule the user's tasks on the clusters.
- *Hadoop Map Reduce* – The implementation of the 'Map-Reduce' programming model for the purpose of processing large amounts of data.

6.2. SAP-HANA

SAP HANA is a column-oriented RDBMS developed and marketed by SAP. It was previously called "SAP High-Performance Analytic Appliance". It basically uses the concept of processing blocks of data using parallel architecture and algorithms to improve performance and get efficient and accurate results.

6.3. Spark

Spark systems provide a scalable computing platform and makes use of the concept of in-memory computing. In-memory computing improves performance. This is an open-source environment and is basically designed for applications like machine-learning and natural language processing. Spark applications run on Apache Mesos and they can co-exist along with Hadoop applications. It uses the programming model of Transform-Action that is similar to Map-Reduce used in Hadoop. It is based on scala and can support data processing in both iterative and distributed environments

7. References

- [1] Hiba Asri, Hajar Mousannif, Hassan Al Moatassime, Thomas Noel, 'Big Data in healthcare: Challenges and Opportunities', IEEE, 2015.
- [2] J.Archenaa and E.A.Mary Anita, 'A Survey Of Big Data Analytics in Healthcare and Government', ISBCC'15, Elsevier, pp. 408-413, 2015.
- [3] Kyoungyoung Jee, Gang-Hoon Kim, 'Potentiality of Big Data in the Medical Sector: Focus on How to Reshape the Healthcare System', Health Informatics Research, pp. 79-85, June 2013.
- [4] Nina S. Godbole, John Lamb, 'Using Data Science & Big Data Analytics to Make Healthcare Green', IEEE, 2015.
- [5] Jigna Ashish Patel, Priyanka Sharma, 'Big data for Better Health Planning', IEEE International Conference on Advances in Engineering & Technology Research, 2014.