

A Review on Issues related to Big Data Analytics in Health Care

B.Swapna^{#1}, Anuradha.T^{#2}

Department of Information Technology
Velagapudi Ramakrishna Siddhartha Engineering College
Vijayawada, India

Abstract-- with the massive growth in the information Technology, data is being collected by organizations from different sources having different formats. Every organization whether it may be a small one or big one, has lots of relevant data accumulated. Though analysing this big data is very essential for the success of the organization, so many issues to be considered because of the size and diversity of data. This paper discusses various issues related to big data analytics in Health Care.

Keywords — Big data analytics, Big data architecture, Data Mining, EMR, Health Care

I. INTRODUCTION

Nowadays a large volume of data is being accumulated with every organization which is being collected for so many years and it is also becoming easy to capture and access data through third parties (1). Different varieties of data is available like personal details and information related to users, shopping or purchasing data related to different clients, data related to social networking sites, logging information related to web or data generated by different sensors which are of different formats. The proper analysis of this data is effecting different organizations like marketing (2,3), business(4), advertising, social networking(5), health care and many other fields (6,7,8). Over the years analysis is done using data mining algorithms, parallel and distributed mining algorithms (9,10,11).

The health care data volume is increasing rapidly and this increase is expected to be enormous in the years ahead with the development of Electronic Health Records (12). With the growth of Automated Disease Prediction systems and Mobile Health Applications, there is a need to digitise patient related clinical data. However, the traditional data managing ways using relational data and data mining are not capable for health care data as the medical data will be in different formats like structured, semi structure and unstructured(13). The data will be of different types like text files, images, sounds and streaming data. Big data analytics which deals with large, diversified data is appropriate for health care analytics(14).

II. THE 3 V'S OF HEALTH CARE DATA

Data Analytics in healthcare is a merging of engineering technology with clinical observations and innovations. As the healthcare industry is continuously generating large amounts of data in different forms, it is almost impossible to manage this data over soft or hard copy formats. The current trend favours digitization and analysis of this large amount of data. Analysing such data supports a wide range of healthcare functions to improve services and tackle problems in healthcare sector. Based on the size and diversity of data, the characteristics of Big Data (15) are very apt for health care organizations.

- Volume-The size of the data determines whether it can be considered as Big Data. Healthcare data contains lots of new as well as historical data.
- Variety –The type and nature of data. Variety refers to structured and unstructured data, textual and multimedia data. Electronic Medical data contains textual lab reports, hand written prescriptions, images from scans or ECG reports, X-rays etc.
- Velocity – The speed at which data is generated. The data from the devices connected to a patient in intensive or critical care units generates continuous stream of data.

III. DATA ANALYTICS IN HEALTHCARE

Data analytics in health care can be used to improve the standards in the following areas(16,17):

- Public Health: By analysing disease patterns and recording disease outbreaks, public health issues can be improved with analytics approach. Large amount of data can help determine needs, offer required services and predict and prevent the future crises to benefit the population.
- Electronic Medical Record or EMR: Every patient has his/her electronic

health Record which contains demographics, allergies, lab reports. An EMR contains the standard (structured and unstructured) medical data that can be evaluated with the data analytic approach to predict patients at risk and provide them effective care by providing warnings or reminders about when a patient should take a certain lab test.

- Patient Profile Analytics: Advanced analytics can be applied to patients' profile for identifying individuals who could benefit from proactive approach. This may include lifestyle changes.
- Genomic: This approach can be effectively included in genomic analytics to make this approach a part of regular medical care decision process.
- Fraud Analysis: This data analytics approach helps analyse greater number of claim requests to curtail down fraud cases. An effective analysis can help reduce fraud, waste and abuse.
- Safety Monitoring: Data analytics can also be used to analyse real time large volumes of brisk data in hospitals. The approach may help in the safety monitoring and negative event prediction.

IV. BIG DATA ARCHITECTURE IN HEALTH CARE

The architectural framework for Health Care Big data shown in Fig.1 is similar to that of a conventional health analytics project (12, 18). The main difference lies in the way how processing of data is done. In traditional system, analysis is done with an analytical tool installed on a standalone computing machine. Because of the volume of big data, processing is done at different nodes. The availability of open source platforms such as Hadoop/Mapreduce on cloud further supported the concept of Big Data analytics in health care. The algorithms used for analysis of health care data in traditional systems and big data system is almost same but the complexity lies in the data itself.

Big data in health care comes from different sources namely internal sources like electronic health records and external sources like laboratories, pharmacies, insurance companies and will be in different formats (they can be text files, .csv files or relational tables etc.). Data can reside at different geographic locations and it

may reside on different health care providers' local systems. Different sources of data include:

1. Data from different mobile health apps or from health plan websites
2. Data from remote sensors or meters readings
3. Data generated from health insurance claims and other billing records
4. Different formats of data generated from X-rays, retinal scans and other reports in image formats, biometric data like finger prints, genetics and bp, pulse meter readings etc.
5. Data generated by the humans like doctor's notes, prescriptions, emails or other paper documents

For the purpose of big data analytics, this raw data which is in structured, semi structured and unstructured formats need to be collected from different sources and then it should be pre-processed or cleaned using ETL tools and depending on the data, it can be in different formats. The transformed data is stored and analysed using different tools and the required analysis reports will be generated.

V. CHALLENGES IN HEALTH CARE DATA ANALYTICS

The advancements in the information technology along with the increase in the storage and processing power of the systems has paved the path for electronically storing, maintaining and moving the health care data in a fraction of seconds across the globe. This made the electronic systems to provide quality of services but there are some challenges in health care data analytics(19).

Security and Privacy

Health information needs to be shared among different researchers involved in doing the research for finding the causes, treatments and remedies, among different health care service providers to analyse and better understand about the latest services to be provided. In this process, data may be collected from different sources like mobile devices or different local systems and communicated through network, care should be taken so that data will not go to unauthorized users and as the data also involves patient's personal data, care should be taken to provide privacy.

Data Management

The tremendous increase in the size of health care data is another issue which need to be dealt. The issue is not just about the size, but also about the tremendous increase in the speed in which this data is being generated. And also the data is complex, being structured, semi structured and unstructured. The new technological developments in health data capturing such as sensors, mobile devices and others are producing complex varieties of data and it is difficult to store and process using traditional database management systems.

Collecting the data from different devices, storing, searching and analysing health care data will definitely benefit the public as well as healthcare industry. But it requires powerful analytical algorithms and smarter decision support systems and computing which can deal with complex medical data.

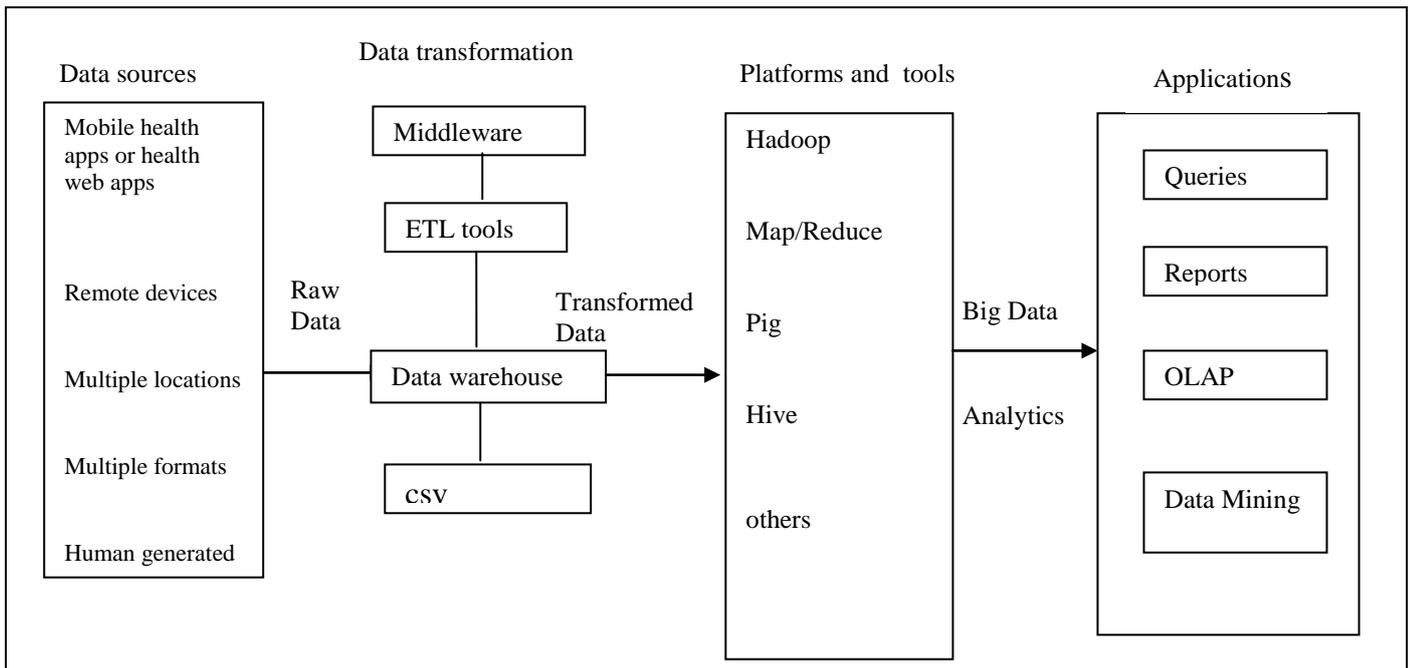


Fig 1: Architecture of Health Care Data Analytics

VI. CONCLUSIONS

In this review paper issues like how the characteristics of big data are suitable to health care, applications of big data analytics in health care, architecture of big data with respect to health care, challenges in health care analytics were discussed.

REFERENCES

[1] Richa Gupta, "Journey from Data Mining to Web Mining to Big Data", *International Journal of Computer Trends and Technology (IJCTT)*, volume 10 number 1, Apr 2014.

[2] M.Suman, T.Anuradha, K.Manasa Veena, "Direct Marketing With the Application of Data Mining", *International Journal of Engineering Research and Applications (IJERA)* Vol. 2, Issue 1, pp.041-043, Jan-Feb 2012

[3] K. Prathyusha, T. Anuradha, R.Sai Nikitha and K. Meghana, "Detecting Frauds In Online Auction System", *International Journal of Advanced Research in Computer Science and Software Engineering* Volume 3, Issue 4, pp.42-47, April 2013.

[4] T.Lakshmi Surekha, P. Rama devi, " Profitable Association Rule Mining Using Weights", *International Journal of Computer Trends and Technology" IJCTT*, May 2016.

[5] G. Ravali1, P.Swathi Mouluka, A.A parna, J.Abhinaya, T.Anuradha, "Analysing You Tube Data using K-Means Clustering", *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)* Volume 6, Issue 2, pp.62-66, March - April 2017.

[6] Vineela.A, T.Anuradha, "Recent Trends In Data Mining For Crime Detection", *International Journal of Applied Engineering Research* Volume 9, Number 24, pp. 24611-24616, 2014.

[7] T Anuradha, Sri Sai A Krishna, K Satyatej, Naga Anil G Kumar, "Frequent Pattern Mining For Efficient Library Management", *International Journal on Computer Science and Engineering*, Vol3 Issue 11 pp.3582-3586

- [8] Anuradha.Tadiparthi, Satya Prasad.R, Tirumala Rao S.N "Identifying Week Subjects Using Association Rule Mining", *International Journal of Scientific & Engineering Research* Volume 2, Issue 11, pp.1-3, November-2011.
- [9] T.Anuradha, M.Kranthi, "Recent Trends in Parallel Clustering and Classification Data Mining", *GJCAT*, Vol 1 (4), PP. 617-619, 2011.
- [10] T.Anu Radha ,P.Lavanya, "Recent Trends in Parallel and Distributed Apriori Algorithm", *International Journal of Engineering Research and Applications*, Vol. 1, Issue 4, pp.1820-1822
- [11] Anuradha.T, Satya Prasad.R, "Parallelizing Apriori on Hyper-Threaded Multi-Core Processor", *International Journal of Advanced Research in Computer Science and Software Engineering* Volume 3, Issue 6,pp. 1072-1082, , June 2013 .
- [12] Raghupathi, Wullianallur, and Viju Raghupathi. "Big data analytics in healthcare: promise and potential." *Health Information Science and Systems* 2.1 (2014): 3.
- [13] Frost & Sullivan, "Drowning in Big Data? Reducing Information Technology Complexities and Costs for Healthcare Organizations". <http://www.emc.com/collateral/analyst-reports/frost-sullivan-educing-information-chnologycomplexities-ar.pdf>
- [14] J.Sun and C. K. Reddy, "Big Data Analytics for Healthcare" Tutorial presentation at the *SIAM International Conference on Data Mining*, Austin, TX, 2013.
- [15] Anuradha.T Meruva Srinadh, Majeti Naveen "Evolution and Challenges of BigData Syatems", *International Journal of Advance Engineering and Research Development* ,vol 1 issue 11 pp 276-282, November 2014
- [16] big-data-examples-in healthcare
Available: <http://www.datapine.com/blog/>
- [17] Burghard C, "Big Data and Analytics Key to Accountable Care Success". *IDC Health Insights*; 2012
- [18] Sawant, Nitin, and Himanshu Shah."Big Data Application Architecture." *Big Data Application Architecture Q & A. Apress*, 9-28, 2013.
- [19]big-data-in-healthcare-made simple
Available: <https://www.healthcatalyst.com/>