International Journal of Latest Trends in Engineering and Technology
Special Issue SACAIM 2016, pp. 206-212
e-ISSN:2278-621X

APPLICATION OF BIG DATA IN BIOINFORMATICS – A SURVEY

Reshma Martiz[1], Supaksha M A[2] and Hemalatha N[3]

Abstract— From past few years big data has made a great booming effect on bioinformatics. It is vast and miscellaneous research field. Researchers from all over the world has made several attempts to understand the field of bioinformatics by analysing its application and tools. These methods can be used to handle big data using various and allocated computing technologies. In this review paper we address a few applications of big data and gives us an overview of its present and helps us to understand the future research opportunities.

Keyword- big data, bioinformatics, genomics, DNA, proteomics.

I. INTRODUCTION

Big data the name itself suggests peaks of data it is an extremely large data sets that may be analysed computationally to reveal patterns, trends, and it is heterogeneous and not specific to a particular field It can also be defined as the high-volume, high-velocity or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation. There is no exact rule about what size a database should be in order for the data to be considered "big." Instead what defines big data is the need for new techniques and tools in order to be able to process it. In order to use big data we need programs which span multiple physical or virtual machines working together. As we enter into the information age data are being generated by variety of sources other than people and servers such as sensors embedded into phones and wearable devices, video surveillance cameras, set-top boxes and so on. High performance technologies are used in scientific research, such as fast data capturing tools and very high resolution satellite data recording.

The world of big data is changing dramatically before our eyes, from the increase in big data growth to the way in which it is structured and used. The trend of big data growth presents enormous challenges, but it also presents incredible business opportunities. Taking growth chart we can see is the rapid growth of Big Data. Keeping that in mind the annual growth of data generation may reach 44 trillion zettabyte by the year 2020.

Big data exists in a wide variety of data intensive areas such as atmospheric science, genomics, research, astronomical studies and network traffic monitor. Huge data is created every day. By the interaction of trillions of people using computer, GPRS, sensors etc. Due to

[1] *Aloysius Institute of Management and Information Technology, Mangalore, Karnataka, India.*
[2] *Aloysius Institute of Management and Information Technology, Mangalore, Karnataka, India.*
[3] *Aloysius Institute of Management and Information Technology, Mangalore, Karnataka, India.*

the tremendous amount of data generated daily from business, research and science, big data is everywhere and represent huge opportunity to those who can use it effectively.

Accuracy in big data may lead to more confident decision making and better decisions can result in greater effective productivity cost reduction and reduced risk.Big Data has the potential to help companies improve operations and make faster, more intelligent decisions. This data, when captured, formatted, manipulated, stored, and analysed can help a company to gain useful insight to increase revenues, get or retain customers, and improve operations.

II. APPLICATIONSOF BIG DATA

A. Biology

From past few years the field of life science has seen a rapid change ingenomics, DNA sequences, gene expression, proteomics and metabolomics etc. There are two main brisk vicissitudes going in the field of life science one is the capabilities of the computers and software tools are increasing from terabytes to Zettabytes and beyond that and the other is the progression in molecular biology experiments producing enormous amount of data related to genome and RNA sequence, protein and metabolism, protein-protein and protein-DNA interaction, gene expression, 3D structure of protein molecules and many more. Life science and computers have become balancing bridge to each other to form different branches of science like the combination of versatile knowledge that caused the dawn of big data in biology, bioinformatics, biostatistics and many more branches.

The low cost of data generationin the field of biology is leading us to the big data. Big data sources are no longer limited to particular field or search-engine or indexes. With digitization with modern technology and availability of high quantity of devices at lower costs. The raise of data volume in research is taking a big change in big data era

The tendency in increasing data volume is reinforced by lessening computing cost and increasing analytics amount with growing technologies. Biologists no longer use traditional laboratories to find out the results, rather they rely on huge and continuously growing of technologiesfor capturingsequence’s thus making it cheaperandmoreeffective.As sequencing technology has improved, the volume of sequence data being produced has begun to outstrip the competences of computer hardware employing conventional methods for analysing. Datasuch as automated genome sequencers, DNA computing and many more giving rise to this new era of big data in bioinformatics.

Talking about bioinformatics in big data it has become a dynamic site of research area which is capable of amassing the benefits from the Big Data. The data can be in many forms which may include information about genomic sequences, molecular pathways, of different populations of people. If we can understand how to handle the sophisticatedinformation tools and techniques for analysing big data. It promises to mould massive mounts of information into a better understanding of the basic biological mechanisms and how the results can be applied in. Big data is one of the biggest trait of biological studies, researchers are capable of generating petabyte of data within hours and it possess a great impact on the bioinformatics like sequencing data, genomic sequence, protein sequence, DNA computing etc.

The developing technology has made data entry far easierthan first with this emerges a great challenges for researchers. Data have flourished so large that to obtain and analysing a data with old technique has become challenging. More over these data can afford an exciting opportunity to recognise large scale patterns and make predictions like protein structure or genomics and many more. The Big Data Analytics in Bioinformatics blends the fields of

biology, technology, and medicine in order to present a complete study on the present information. Toovercome the drawbacks, there are special methodologies and tools to manage massive and complex data sets which allows us too quickly and effectively hitch the power of Big Data to make ground-breaking biological discoveries, carry out translational medical research, and implement personalized genomic medicine.[1][2]

B. Personalized Medicine

Currently researchers are trying hard to meet the present demands in the field of personalized medicine by studying various ways to collaborate and coordinate medicine with new technologies. Tool are been developed to reduce cost to improve patient's safety and healthcare quality. By Changing the way to design and mange treatment trials and also by using big data to bring personalized medicine to drug trials and research, can potentially reduce costs, allow the right drugs to be work faster, and improve outcomes at lower costs. It also helps in faster drug development with easier margins. Big data in Personalized Medicine is one the biggest game changer for pharmaceutical companies. Manufacturers are looking forward to develop high-value, cost effective and targeted drugs. It has long been maintained that personalized medicine offers the chance of better health with less harm. Making assumptions using a single mind is no longer enough. So there is an urgent emerge of combining both the IT and biology together to find an innovative answer(Figure 1).The volume of data available today is extraordinary, these changes have shaped an effective outbreak of mining data on genomics, proteomics, and metabolomics alongside clinical-trial data and real-world clinical data will allow researchers to more completely understand the particular structure of a disease. This awareness will lead to faster and more accurate identification and validation of targets and biomarkers for proposing personalized therapies [3].

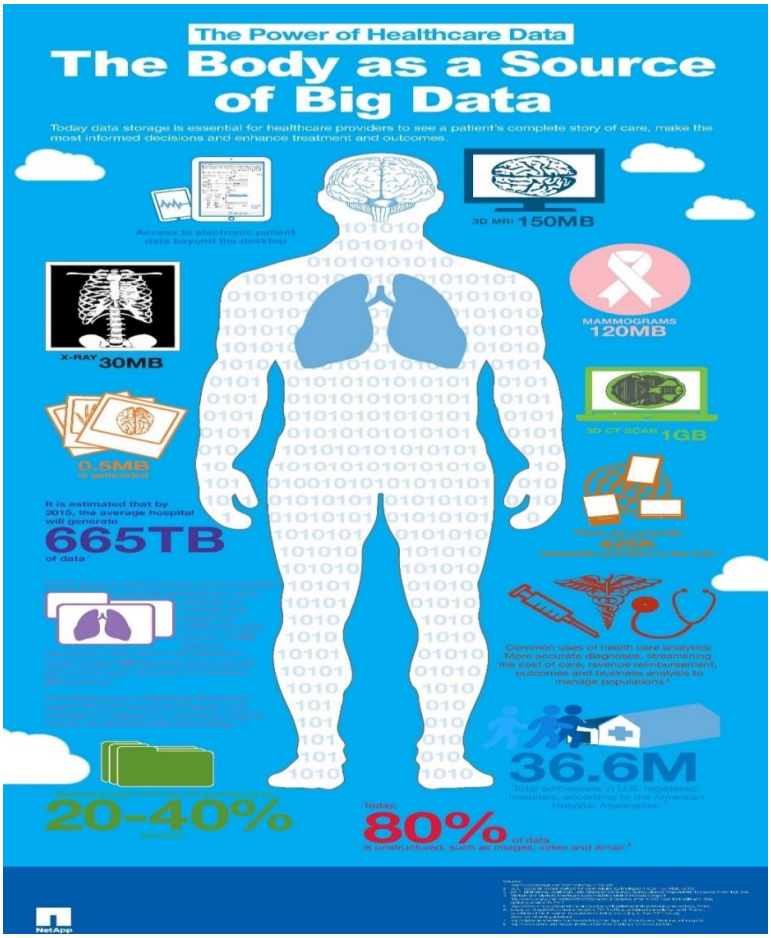

Figure 1-Depicts Big data in personalized medicine

C. Gene Sequences

Researchers are getting attracted by new outlooks on the human genome, and the progression made in big data analytics. For many years genes have been studied, mapped and experimented in gene sequencing. Perhaps the highest achievement was accomplishment of the Human Genome Project in the early 2000s. According to biologist by 2025, 100 million to 2 billion human genomes could have been sequenced where simultaneously the data storage demands also increases from 2–40 petabyte, because the number of data that must be stored for a single genome are 30 times larger than the size of the genome itself. But identifying how human genetics work has required more serious study with more supplies. Only recently some major changeshas taken place to handle the volume of data and the speed of analysis. Now scientists will be able to look more closely at human genes, and much of this progress comes as they apply big data analytics to these issues[4][5].

D. Preventative medicine

System biology uses computational and mathematical modelling to solve complex biological problem. It is now making a great deal in healthcare and medicine with the help of digital revolution. The main goal is to provide better foundation to produce and develop preventative, predictive in the field of medical science. Recently many advance computer

support systems which helps us to improve protect, promote, and maintain health and well-being and to prevent disease, disability and death.

Using big data in field of preventative medicine, we can improve the health of patients and give a better diagnose while treating the disease. As the role of Big data comes, more and more information from all around the world can be balanced. As the prototypes are being made with the help of large collection of data using big data technique, it is easy to measure the outcome.[6]

E. Healthcare

Due to the massive collection of information, the healthcare industry has entered into the new era to generate large amount of data by keeping its record and maintain its requirement in patient's health and care. Due to this eruption of information big data analysis in healthcare has been the debated topic in growing healthcare information system today. Using big data it is easy to identify the fraud claims, it helps in transformation of medical claims payment system. Using this technique it is easy to treat a wide range of conditions such as diabetes, chronicle disorders, heart diseases, various cancer conditions. It also helps to improve identification and measurement of quality metrics. This analytical technique provides vast opportunities in investigation of clinical trials of dataset. This field provides a correct understanding of dataset. For example, a patient can have a combination of characteristics that help to identify form other by implementing data from similar patient it will be easy to identify the disease and helps to provide the best treatment, for better health. This analytical technique also help in cost reducing treatment for better health of patient. Even electronic records and tools are available to identify the condition of patients which helps to give the best care available. [7][8]

F. Industry

Bioinformatics involves the application of data-rich computational and informatics methods to support the scientific study of complex biological problems. Recent technological and intellectual advances in bioinformatics are transforming the way biological research is conducted, placing bioinformatics at the forefront of the information-intensive approach to scientific investigation. As a result, biological research is currently experiencing explosive growth in academic, industry, and government sectors.This visualisation creates the more need for information along with the data mining, digital libraries, modelling and simulation, and other information. The prominent field of the bioinformatics visualisation indicates the design of visual image and the executionof effective software tools that provide an accurate and deep understanding into complex biological data. The flow of big data can be explained with the diagram given in Figure 2

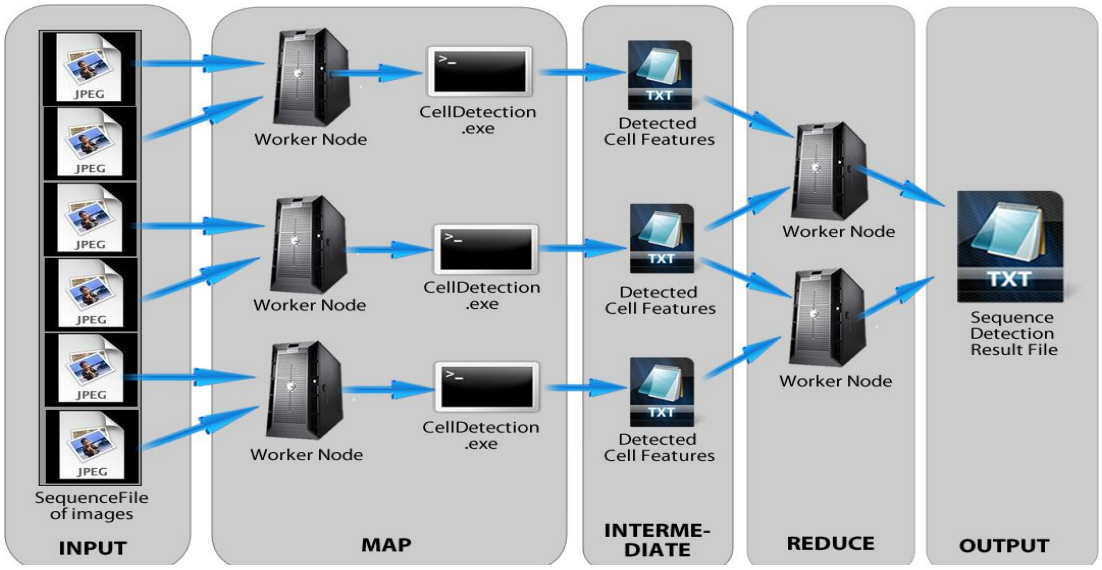

Figure 2-Depicts Big data in Industry

There are some characteristics of bioinformatics that put forward many challenges to visualisation researches. It contain large amount of data like high profiling technologies functional genomics, proteomics, and metabolomics, combined with an emphasis on online heterogeneous data. Data integration consists in providing a uniform view of a set of heterogeneous data sources. This allows users to define their queries without any knowledge on the heterogeneous sources. Data integration systems use the mediation architecture to provide integrated access to multiple data sources. Complex exploratory tasks: recent biological research mainly focuses on statistical testing of specific theory , many areas such as systems biology are increasingly bring an open-ended exploratory approach to theory, generation and large scale data analysis to understand exceedingly complex biological phenomena.[8][9]

They have been successfully used in many domains of bioinformatics, such as bio-molecular structure visualization, expression profile analysis, sequence analysis and annotation, genome visualization, molecular pathways and hierarchical biological data. The current applications of information visualization methods in bioinformatics for visualizing different types of biological data, such as from genomics, proteomics, expression profiling and structural studies.[9][10]

III .CONCLUSION

In this review paper we have discussed recent application in bioinformatics and data in terms of volume. Big data is a promising research area, still in its formative years with the initiation of new high information and cost reducing data capturing tools. The big data analytics techniques are required to solve the problems in bioinformatics such as storage of vast information generated by analysing the structure of data. In this paper we have to investigate the role of big data in various fields. Though only few topics have beendiscussed here there are many areas where big data tools and techniques can be applied,in future years it is going to be a boon for researchers to overcome all the hurdles and use big data techniques in very aspects of life science.

REFERENCE

[1] Big Biological Data: Challenges and Opportunities Yixue Li, , Luonan Chen
[2] Wang, Baoying, ed. Big Data Analytics in Bioinformatics and Healthcare. IGI Global, 2014.
[3] Marrying Big Data with Personalized Medicine Data analytic strategies can help companies capitalize on personalized medicineBy Michael Kuchenreuther, Jill E. Sackman, PhD.
[4] Big Data Analytics Alters How We Study the Human Genome By Jonathan Buckley.
[5] Genome researchers raise alarm over big data By Erika Check Hayden.
[6] http:www.mckinsey.com/industries/pharmaceuticals-and-medical-products/our-insights/the-role-of-big-data-in-medicine.
[7] A Survey On Big Data Analytics In Health Care.
[8] Transforming Health Care Through Big Data.
[9] Information Visualization Techniques in Bioinformatics during the Postgenomic EraYing Tao, Yang Liu, Carol Friedman, and Yves A. Lussier.
[10] Introduction to this Special Issue of Information Visualization