

Chapter 6

Data Mining, Big Data, Data Analytics: Big Data Analytics in Bioinformatics

Priya P. Panigrahi

Jaypee University of Information Technology, India

Tiratha Raj Singh

Jaypee University of Information Technology, India

ABSTRACT

In this digital and computing world, data formation and collection rate are growing very rapidly. With these improved proficiencies of data storage and fast computation along with the real-time distribution of data through the internet, the usual everyday ingestion of data is mounting exponentially. With the continuous advancement in data storage and accessibility of smart devices, the impact of big data will continue to develop. This chapter provides the fundamental concepts of big data, its benefits, probable pitfalls, big data analytics and its impact in Bioinformatics. With the generation of the deluge of biological data through next generation sequencing projects, there is a need to handle this data through big data techniques. The chapter also presents a discussion of the tools for analytics, development of a novel data life cycle on big data, details of the problems and challenges connected with big data with special relevance to bioinformatics.

INTRODUCTION

Big data has undoubtedly gained much attention in every sector like science, IT, social media, etc., in the 21st century. Myriad technological revolutions are pouring the intensification of data and data gathering. In recent years we have observed a histrionic growth in data availability. For example; the number of web pages indexed by Google, which was nearly one million in 1999, have exceeded at 4.73 billion pages in 2015, and its enlargement is speeded up by the existence of the social networks (Che *et al.*, 2013, Grobelnik, 2012, Worldometers, 2014).

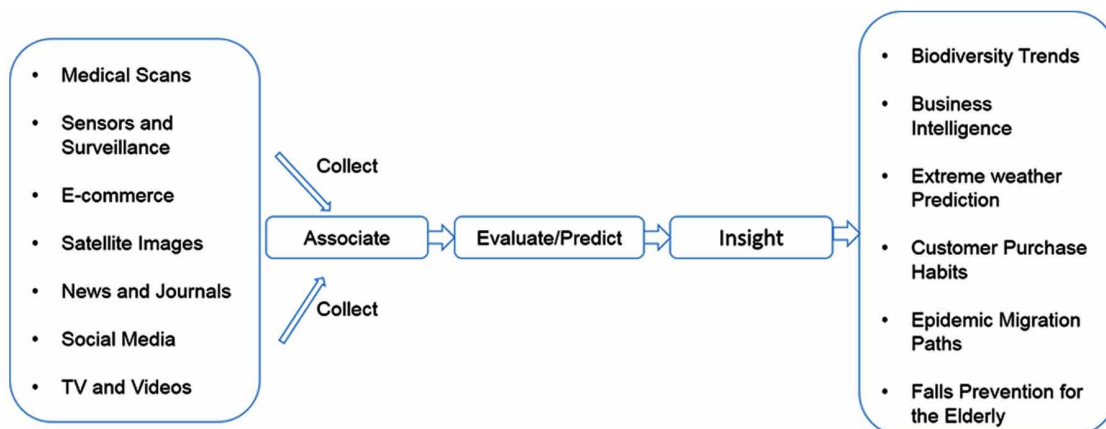
DOI: 10.4018/978-1-5225-7501-6.ch006

Why 'Big Data' Is Essential

- Numerous diverse big data programs launched.
- Increased use of sensors in all sectors like traffic patterns, purchasing behaviors, real-time inventory management, etc.
- Supermarkets handle approximately 1 million consumer transactions every single hour, which is imported into databases estimated to have nearly three petabytes of data.
- 72 hours of video are added to YouTube every minute.
- There are approximately 217 new mobile internet users every minute.
- Facebook handles 40 to 50 billion photographs from its user base.
- Twitter users send more than 150 million tweets per day.
- Biomedical computation decoding human Genome and personalized medicine.
- Social science revolution, etc. (Shaw, 2014).

As a result of this huge amount of data 'big data' has become a modern area of potential investment. According to The McKinsey Global Institute (2012), "Big data refers to data sets whose dimension is beyond the capability of usual database software tools to capture, store, manage and analyze" (Manyika, *et al.*, 2012). In Gartner June 2012 issue, Beyer and Laney stated: "Big data are high volume, high velocity, and high variety information resources that necessitate innovative procedures to enable heightened outcomes, insight discovery, and process optimization" (Beyer & Laney, 2012). Therefore, big data needs different methods, tools, and architectures to decipher novel problems and deep-rooted problems in an improved way. Some crucial factors for the evolution of big data are; accessibility of data, enlargement of storage capabilities, and enlargement of processing power, etc. Firms in most sectors have a minimum of 100 terabytes of stored data, and several have more than a petabyte. The size of big data is so vast and multifaceted that ordinary data handling applications are insufficient. The challenges comprise search, capture, data curation, exploration, storing, allocation, conception, and data privacy (Khan *et al.*, 2014). Figure 1 describes the basic workflow of big data architecture. Precision in big data can lead to more poised results and ultimately improved decisions that lead to better operational competence, cost

Figure 1. Big data facilitates institutions to collect, accumulate, and manipulate massive quantity of data at the right speed and right time



14 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/data-mining-big-data-data-analytics/217824

Related Content

The Development, Testing, and Deployment of a Web Services Infrastructure for Distributed Healthcare Delivery, Research, and Training

Ghita Kouadri Mostefaoui (2009). *Managing Web Service Quality: Measuring Outcomes and Effectiveness* (pp. 1-22).

www.irma-international.org/chapter/development-testing-deployment-web-services/26072

Adaptation Space: A Design Framework for Adaptive Web Services

Henrique Paques, Ling Liu and Calton Pu (2004). *International Journal of Web Services Research* (pp. 1-24).

www.irma-international.org/article/adaptation-space-design-framework-adaptive/3042

Semantic Annotation of Geospatial RESTful Services Using External Resources

Victor Saquicela, Luis. M. Vilches-Blázquez and Oscar Corcho (2012). *Discovery of Geospatial Resources: Methodologies, Technologies, and Emergent Applications* (pp. 156-171).

www.irma-international.org/chapter/semantic-annotation-geospatial-restful-services/65113

Model-Based Development of Web Services Descriptions Enabling a Precise Matching Concept

Jan Hendrik Hausmann, Reiko Heckel and Marc Lohmann (2005). *International Journal of Web Services Research* (pp. 67-84).

www.irma-international.org/article/model-based-development-web-services/3060

A Qualitative Approach to Understand Consumer Groups and Decision-Making Process in Online Group Buying: An Exploratory Study

Lin Xiao and Chuanmin Mi (2019). *International Journal of Web Services Research* (pp. 24-46).

www.irma-international.org/article/a-qualitative-approach-to-understand-consumer-groups-and-decision-making-process-in-online-group-buying/226288