Chapter 52
Data Science and Computational Biology

Singaraju Jyothi
Sri Padmavati Mahila University, India

Bhargavi P
Sri Padmavati Mahila University, India

ABSTRACT

Data Science and Computational biology is an interdisciplinary program that brings together the domain specific knowledge of science and engineering with relevant areas of computing and bioinformatics. Data science has the potential to revolutionise healthcare, and respond to the increasing volume and complexity in biomedical and bioinformatics data. From genomics to clinical records, from imaging to mobile health and personalised medicine, the data volume in biomedical research presents urgent challenges for computer science. This chapter elevates the researchers in what way data science play important role in Computational Biology such as Bio-molecular Computation, Computational Photonics, Medical Imaging, Scientific Computing, Structural Biology, Bioinformatics and Bio-Computing etc. Big data analytics of biological data bases, high performance computing in large sequence of genome database and Scientific Visualization are also discussed in this chapter.

INTRODUCTION

The term "Data Science" has emerged only recently to specifically designate a new profession that is expected to make sense of the vast stores of big data. Data Science draws scientific inquiry from a broad range of subject areas such as statistics, mathematics, computer science, machine learning and optimization, signal processing, information retrieval, databases, cloud computing, computer vision, natural language processing and etc. Data Science is on the essence of deriving valuable insights from data. It is emerging to meet the challenges of processing very large datasets, i.e. Big Data, with the explosion of new data continuously generated from various channels such as smart devices, web, mobile and social

DOI: 10.4018/978-1-5225-8903-7.ch052

media. At a high level, data science is a set of fundamental principles that support and guide the principled extraction of information and knowledge from data. Possibly the most closely related concept to data science is data mining-the actual extraction of knowledge from data via technologies.

A data science perspective provides practitioners with structure and principles, which give the data scientist a framework to systematically treat problems of extracting useful knowledge from data. The past of data science was all about descriptive analytics, or describing what has already taken place. But the future of data science will hinge on advanced analytics, specifically using predictive analytics and real time analytics. In the future, "the tools are going to de-emphasize the mechanics of doing machine learning.

Underlying the extensive collection of techniques for mining data is a much smaller set of fundamental concepts comprising data science. In order for data science to flourish as a field, rather than to drown in the flood of popular attention, data science is thought as beyond the algorithms, techniques, and tools in common use. The core principles and concepts that underlie the techniques and also the systematic thinking that foster success in data-driven decision making. These data science concepts are general and very broadly applicable. Data science supports data-driven decision making and sometimes allows making decisions automatically at massive scale and depends upon technologies for ''big data'' storage and engineering. However, the principles of data science are its own and should be considered and discussed explicitly in order for data science to realize its potential.

Computational Biology or Bioinformatics has been defined as the application of mathematical and Computer Science methods to solving problems in Molecular Biology that require large scale data, computation, and analysis. As expected, Molecular Biology databases play an essential role in Computational Biology research and development. Computational biology is part of a larger revolution that will affect how all of science is conducted. This larger revolution is being driven by the generation and use of information in all forms and in enormous quantities and requires the development of intelligent systems for gathering, storing and accessing information (Clutter, 1996).

Extracting knowledge from data is a defining challenge of science. Computational genomics has been an important area since the beginning of the Human Genome Project. Today, however, advances in tools and techniques for data generation are rapidly increasing the amount of data available to researchers, particularly in genomics. This increase requires researchers to rely ever more heavily on computational and data science tools for the storage, management, analysis, and visualization of data. These efforts support research and development of transformative approaches and tools that maximize the integration of Big Data (like genomics data) and data science into biomedical research.

1. DATA SCIENCE

Day to day the amount of data produced across the world has been exponentially increasing and will continue to grow in the future. At companies across all industries, servers are overflowing with usage logs, message streams, transaction records, sensor data, business operations records and mobile device data. Effectively analyzing these huge collections of data, can create significant value for the world economy by enhancing productivity, increasing efficiency and delivering more value to consumers. Studies estimate that trillions of dollars of value in efficiency improvements and economic growth can be unlocked by extracting actionable knowledge from the deluge of data now being collected in almost every sector of the economy.