Raj Bhatnagar

Professor of Computer Science,
University of Cincinnati, Cincinnati, OH 45221, USA

Topic

Algorithms for Big Data Analytics.

Main Focus

Big data analytics related problems are being encountered in business, science, and engineering. There are various different types of analyses that need to be performed with huge, distributed, and heterogeneous types of databases. A large number of machine learning and data mining algorithms have been developed and investigated during the past few years. Most of them have been investigated in the context of large datasets but much smaller than the giga and tera bytes sizes that are now being generated and retained. Each existing algorithm caters to some specific types of patterns or properties of data, and also works for some specific type of data. All these algorithms have been developed with the assumption that the dataset or the data warehouse is resident at a single computer system. In the big data scenarios, we may have a number of databases, each may reside on a different node of an interconnected set of servers in a Cloud environment, and each is generally too large to be moved to other nodes across the network for processing. So there is need for algorithms that adapt themselves to the constraints of the data, that is, they can process tera-scale data, and work with data wherever it resides instead of demanding data to be located for their convenience and efficiency. We present in this tutorial all the adaptations and advances that are needed in the analysis algorithms so that they can meet the challenges of the big datasets. Some algorithmic developments of the recent past will be presented, and a lot of remaining challenges will be outlined, each in the context of problem that needs these algorithms.