

Chapter 10

Discovering Knowledge Hidden in Big Data from Machine–Learning Techniques

Adiraju Prashantha Rao
Anurag Group of Institutions, India

ABSTRACT

As the speed of information growth exceeds in this new century, excessive data is making great troubles to human beings. However, there are so much potential and highly useful values hidden in the huge volume of data. Big Data has drawn huge attention from researchers in information sciences, policy and decision makers in governments and enterprises. Data analytic is the science of examining raw data with the purpose of drawing conclusions about that information. Data analytics is about discovering knowledge from large volumes data and applying it to the business. Machine learning is ideal for exploiting the opportunities hidden in big data. This chapter able to discover and display the patterns buried in the data using machine learning.

INTRODUCTION

Big Data has been one of the current and future research frontiers. Big Data analytics is a broad topic, fed by a lot of other top technology trends. David Cearley (2016), vice president and Garner Fellow at Gartner Group, shared his thoughts with Information Management on the Top 10 Strategic Technology Trends that will impact IT leaders and data analytics in 2016. He noted that technology based companies such as Google, Yahoo, Face Book, Microsoft, Amazon have been collecting and maintaining data that is measured in terms of Exabyte propositions or larger. Human-sourced information is now almost entirely digitized and stored everywhere from personal computers to social networks. Data are loosely structured and often ungoverned. Internets of Things are derived from the phenomenal growth in the number of sensors and machines, used to measure and record the events and situations in the physical world. The output of these sensors is machine-generated data and from simple sensor records to complex computer logs, it is well structured. As sensors proliferate and data volumes grow, it is becoming an increasingly

DOI: 10.4018/978-1-5225-1877-8.ch010

important component of the information, stored and processed by many businesses. Its well-structured nature is suitable for computer processing, but its size and speed is beyond traditional approaches. The explosion of data sources and growing complexity of information makes manual classification and analysis, infeasible and uneconomic. “Deep Neural Nets (DNNs) automate these tasks and make it possible, to address key challenges related to the information of every trend” (Cearley, 2016). The machine learning techniques give rise to a spectrum of smart implementation, which explores automated agents to perform all activities without human intervention. Big data refers to large volumes of data structured and unstructured, gathered from various servers and databases. Mining of information is required to extract necessary information which represents the general group of problems and techniques used for application domains that collect and maintain enormous volumes of raw data for specific domain analysis. This means big data is a collection of very huge data sets with great diversity of types which can be very difficult to process using traditional data processing techniques. So data sets are generally known to be Big Data if it is difficult to perform capture, curation, analysis, and visualization on it with current technology. So Big Data requires new forms of processing to enable enhanced decision making, insight discovery and process optimization. In many instances, science is lagging behind the real world in the capability of discovering the valuable knowledge from enormous volumes of data. There is a need to develop and create new technologies and techniques to solve the problems of big data and also enhancing the current technologies. The main contents which are presented in this chapter are, the applications of big data presented, big data definition and its challenges, big data analysis methodology and knowledge discovery from huge volume presented. Finally, the chapter concludes with its future scope.

BIG DATA APPLICATIONS

Patient Data Sensing and Clinical Records

Initially, supporting increasing potential patients in hospitals or care centers requires remote monitoring of patients as a solution, which in turn leads to difficulty in handling challenges of big data i.e. volume, velocity, variety, veracity etc. Also the full cycle of this huge data (i.e. capturing, gathering, clearing, transforming, formats, storing, analyzing and visualizing) for some patients should be covered in real time. Another data source that caters for data diversity and variety is the patient’s clinical records. In all cases, data is made available for access, for all teams of doctors, nurses, administrators and social agents. So the technologies other than Hadoop like batch processing should be considered as better solutions.

Genomics

The genomics is defined as large volume of data (nearly 100 gigabytes) associated to sequencing of DNA from different biological data sources. In case of personalized genomics, number of people wants their DNA sequenced for diagnostic and prognostics purposes, leading to hundreds of thousands of peta bytes of data. Raw DNA sequences are annotated, (making intelligible and rich genomics information) further increasing the data volumes, leading to the development of efficient compression algorithms for sequencing data. The main problems encountered in this process are as follows:

15 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/discovering-knowledge-hidden-in-big-data-from-machine-learning-techniques/173830

Related Content

Multicriteria Evaluation of the Insurance Fund Value at Insurance of Space Projects

(2017). *Multi-Criteria Decision Making for the Management of Complex Systems* (pp. 132-140).

www.irma-international.org/chapter/multicriteria-evaluation-of-the-insurance-fund-value-at-insurance-of-space-projects/180012

In-Store Stimuli and Impulsive Buying Behaviour: Modeling Through Regression Equation

Chandan Parsad, Sanjeev Prashar, T. Sai Vijayand Mukesh Kumar (2018). *International Journal of Strategic Decision Sciences* (pp. 95-112).

www.irma-international.org/article/in-store-stimuli-and-impulsive-buying-behaviour/208681

Implications and Philosophical Requirements of a Comprehensive Dialectical Inquiry System

William Acarand Douglas A. Druckenmiller (2014). *Analytical Approaches to Strategic Decision-Making: Interdisciplinary Considerations* (pp. 1-15).

www.irma-international.org/chapter/implications-and-philosophical-requirements-of-a-comprehensive-dialectical-inquiry-system/102147

Potential Power and Problems in Sentiment Mining of Social Media

John Wang, Qiannong Guand Gang Wang (2013). *International Journal of Strategic Decision Sciences* (pp. 16-26).

www.irma-international.org/article/potential-power-problems-sentiment-mining/78345

Cooperative Decision Support Systems

Pascale Zaraté (2008). *Encyclopedia of Decision Making and Decision Support Technologies* (pp. 109-115).

www.irma-international.org/chapter/cooperative-decision-support-systems/11246