

Chapter 108

Unleashing Artificial Intelligence Onto Big Data: A Review

Rupa Mahanty

Tata Consultancy Services, Australia

Prabhat Kumar Mahanti

University of New Brunswick, Canada

ABSTRACT

We live in an ocean of data. Big Data is characterized by vast amounts of data sized in the order of petabytes or even exabytes. Though Big Data has great potential, Big data by itself has no value unless one can derive meaningful results from it. That is where Artificial Intelligence pitches in. Artificial Intelligence's most common application is about finding patterns in enormous quantities of data. The confluence of Big Data and Artificial Intelligence allows companies to automate and improve complex descriptive, predictive and prescriptive analytical tasks. In other words, Big Data can offer great insights with the help of Artificial Intelligence (AI). Artificial Intelligence can act as a catalyst to derive tangible value from Big data and serve as key to unlocking Big data. This review article focuses on applications of artificial intelligence to Big Data, its Limitations and issues.

INTRODUCTION

In today's digital world, data has grown 'big' – steering in the era of the petabytes and exabytes. Big Data is characterized by astronomical amounts of data being generated continuously by interconnected systems of people, transactions, media, devices (sensors, smartphones, smart meters, cameras and tablet computers) -- click data, audio/speech data, natural language text (in multiple languages), images/video data. The growth of Big data is a result of the wide variety of data and growing channels in today's world. The internet way of things has a significant contribution in the growth of Big data. By 2015, research firm IDC predicts there will be more than 5,300 exabytes of unstructured digital consumer data stored in databases, and we expect a large share of that to be generated by social networks. Facebook ingests

DOI: 10.4018/978-1-5225-7501-6.ch108

approximately 500 times more data each day than the New York Stock Exchange (NYSE). Twitter stores at least 12 times more data each day than the NYSE [Smith, 2014]. The challenge is to analyze the information content in these vast, continuous data streams, use them for descriptive and predictive analytics in various domains and build more robust and intelligent learning systems. With big data benefitting from improved and increased storage capabilities at extremely reasonable prices - the cost of a gigabyte of storage has dropped from approximately \$16 in February

2000 to less than \$0.07 today [SAS, 2012], and with processing technologies specifically designed to handle huge data volumes, thinking moves away from what data/records to keep and store, to muse over the problem of how to make sense and derive logic from these increasing large volumes of data. Yvonne Hofstetter, managing director of Teramark Technologies GmbH, a Germany-based provider of big data technologies and artificial intelligence for the industrial Internet states that the core of big data is the analysis of big data and the inference, which is provided by artificial intelligence (AI) and not storage or retrieval of raw data [O'Dwyer, 2014]. Big Data can offer great insights with the help of Artificial Intelligence (AI).

Artificial Intelligence deals with the study and development of software and machines that can imitate human-like intelligence and it is a branch of computer science that is extremely technical. Artificial intelligence is used in a variety of ways and can be found across a large number of industry sectors- manufacturing, life sciences and healthcare, transportation, and healthcare, finance to name a few. Some examples of its usage are in assembly line robots, advanced toys, online search engines, speech recognition systems, medical research, and marketing. Artificial Intelligence's most common application is about finding patterns in enormous quantities of data. Smaller more homogenous fixed data sets will not serve the purpose as the patterns may not be evident in them. This allows companies to automate and improve complex descriptive, predictive and prescriptive analytical tasks, which would be tremendously labor intensive and time consuming if carried out by humans beings.

The aim of this paper is to explore the opportunities of Big Data focusing on applications of artificial intelligence to Big Data problems. The paper begins with a brief overview of Big data and the characteristics of Big data followed by a sections discussing the application of Artificial Intelligence to Big data and limitations and issues of Big data and Artificial Intelligence.

BIG DATA: AN OVERVIEW

Big Data: Definition

The use of the term "big data" can be traced back to discussions of handling huge groups of datasets in both academia and industry during the 1980s [Yan, 2013]. Michael Cox and David Ellsworth were among the first to use the term big data literally, referring to using larger volumes of scientific data for visualization (the term large data also has been used) [Cox and Ellsworth, 1997].

The first formal, academic definition appears in a paper submitted in July 2000 by Francis Diebold of University of Pennsylvania in his work of econometrics and statistics (2000):

Big Data refers to the explosion in the quantity (and sometimes, quality) of available and potentially relevant data, largely the result of recent and unprecedented advancements in data recording and storage

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/unleashing-artificial-intelligence-onto-big-data/217932

Related Content

Web Service Discovery and Composition for Virtual Enterprises

Jürgen Dorn, Peter Hrastnik and Albert Rainer (2007). *International Journal of Web Services Research* (pp. 23-39).

www.irma-international.org/article/web-service-discovery-composition-virtual/3093

Semantic Enrichment for Geospatial Information in a Tourism Recommender System

Joan de la Flor, Joan Borràs, David Isern, Aida Valls, Antonio Moreno, Antonio Russo, Yolanda Pérez and Salvador Anton-Clavé (2012). *Discovery of Geospatial Resources: Methodologies, Technologies, and Emergent Applications* (pp. 133-155).

www.irma-international.org/chapter/semantic-enrichment-geospatial-information-tourism/65112

Request-Driven Cross-Media Content Adaptation Technique

Mohd Farhan Md Fudzee and Jemal Abawajy (2010). *Developing Advanced Web Services through P2P Computing and Autonomous Agents: Trends and Innovations* (pp. 91-113).

www.irma-international.org/chapter/request-driven-cross-media-content/43649

A Fast Shapelet Discovery Algorithm Based on Important Data Points

Cun Ji, Chao Zhao, Li Pan, Shijun Liu, Chenglei Yang and Lei Wu (2017). *International Journal of Web Services Research* (pp. 67-80).

www.irma-international.org/article/a-fast-shapelet-discovery-algorithm-based-on-important-data-points/181300

Stakeholder Identification and Analysis for Service Lifecycle Management

Wenge Rong, Qinfen Wu, Yuanxin Ouyang, Kecheng Liu and Zhang Xiong (2014). *International Journal of Web Services Research* (pp. 30-56).

www.irma-international.org/article/stakeholder-identification-and-analysis-for-service-lifecycle-management/110873