Big Data Research 2 (2015) 59–64

Contents lists available at ScienceDirect

Big Data Research

www.elsevier.com/locate/bdr

Significance and Challenges of Big Data Research ☆

Xiaolong Jin [a,*], Benjamin W. Wah [a,b], Xueqi Cheng [a], Yuanzhuo Wang [a]

[a] CAS Key Lab of Network Data Science and Technology, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100190, China
[b] The Chinese University of Hong Kong, Shatin, NT, Hong Kong, China

ARTICLE INFO

Article history:
Received 14 November 2014
Accepted 11 January 2015
Available online 26 February 2015

Keywords:
Big data
Data complexity
Computational complexity
System complexity

ABSTRACT

In recent years, the rapid development of Internet, Internet of Things, and Cloud Computing have led to the explosive growth of data in almost every industry and business area. Big data has rapidly developed into a hot topic that attracts extensive attention from academia, industry, and governments around the world. In this position paper, we first briefly introduce the concept of big data, including its definition, features, and value. We then identify from different perspectives the significance and opportunities that big data brings to us. Next, we present representative big data initiatives all over the world. We describe the grand challenges (namely, data complexity, computational complexity, and system complexity), as well as possible solutions to address these challenges. Finally, we conclude the paper by presenting several suggestions on carrying out big data projects.

1. Introduction

In recent years, big data has rapidly developed into a hotspot that attracts great attention from academia, industry, and even governments around the world [1–3]. Nature and Science have published special issues dedicated to discuss the opportunities and challenges brought by big data [4,5]. McKinsey, the well-known management and consulting firm, alleged that big data has penetrated into every area of today's industry and business functions and has become an important factor in production [6]. Using and mining big data heralds a new wave of productivity growth and consumer impetus. O'Reilly Media even asserted that "the future belongs to the companies and people that turn data into products" [7]. Some even say that big data can be regarded the new petroleum that will power the future information economy. In short, the era of big data has already been in the offing.

What is big data? So far, there is no universally accepted definition. In Wikipedia, big data is defined as "an all-encompassing term for any collection of data sets so large and complex that it becomes difficult to process using traditional data processing applications" [8]. From a macro perspective, big data can be regarded as a bond that subtly connects and integrates the physical world, the human society, and cyberspace. Here the physical world has a reflection in cyberspace, embodied as big data, through Internet, the Internet of Things, and other information technologies, while human society generates its big data-based mapping in cyberspace by means of mechanisms like human–computer interfaces, brain–machine interfaces, and mobile Internet [9–11]. In this sense, big data can basically be classified into two categories, namely, data from the physical world, which is usually obtained through sensors, scientific experiments and observations (such as biological data, neural data, astronomical data, and remote sensing data), and data from the human society, which is often acquired from such sources or domains as social networks, Internet, health, finance, economics, and transportation.

Compared to traditional data, the features of big data can be characterized by 5V, namely, huge Volume, high Velocity, high Variety, low Veracity, and high Value. The main difficulty in coping with big data does not only lie in its huge volume, as we may alleviate to some extent this issue by reasonably expanding or extending our computing systems. Actually, the real challenges center around the diversified data types (Variety), timely response requirements (Velocity), and uncertainties in the data (Veracity). Because of the diversified data types, an application often needs to deal with not only traditional structured data, but also semi-structured or unstructured data (including text, images, video, and voice). Timely responses are also challenging because there may not be enough resources to collect, store, and process the big data within a reasonable amount of time. Finally, distinguishing between true and false or reliable and unreliable data is especially challenging, even for the best data cleaning methods to eliminate some inherent unpredictability of data.

Big data is of great value, which is beyond all doubt. From the perspective of the information industry, big data is a strong

☆ This article belongs to Visions on Big Data.

* Corresponding author.
E-mail address: jinxiaolong@ict.ac.cn (X. Jin).

http://dx.doi.org/10.1016/j.bdr.2015.01.006
2214-5796/

impetus to the next generation of IT industry, which is essentially built on the third platform, mainly referring to big data, cloud computing, mobile Internet, and social business. IDC predicted that by 2020 the market size of the third IT platform will reach US$ 5.3 trillion; and from 2013 to 2020, 90% of the growth in the IT industry would be driven by the third IT platform. From the socio-economic point of view, big data is the core connotation and critical support of the so-called second economy, a concept proposed by the American economist W.B. Arthur in 2011 [12], which refers to the economic activities running on processor, connectors, sensors, and executors. It is estimated that by 2030 the size of the second economy will approach that of the first economy (namely, the traditional physical economy). The main support of the second economy is big data, as it is an inexhaustible and constantly enriching resource. In the future, by virtue of big data, the competence under the second economy will no longer be that of labor productivity but of knowledge productivity.

2. Significance of big data

Due to its great value, big data has been essentially changing and transforming the way we live, work, and think [1]. In what follows, we describe in detail the significance of big data in various perspectives.

2.1. Significance to national development

At present, the world has completely entered the era of the information age. The extensive use of Internet, Internet of Things, Cloud Computing, and other emerging IT technologies has made various data sources increasing at an unprecedented rate, while making the structures and types of data increasingly complex. Depth analysis and utilization of big data will play an important role in promoting sustained economic growth of countries and enhance the competitiveness of companies.

In the future, big data will become a new point of economic growth. With big data, companies will upgrade and transform to the mode of Analysis as a Service (AaaS), thereby changing the ecology of the IT and other industries. In this context, the global giants of the IT industry (such as IBM, Google, Microsoft, and Oracle) have already begun their technical development planning in the big data era.

At the national level, the capacity of accumulating, processing, and utilizing vast amounts of data will become a new landmark of a country's strength. The data sovereignty of a country in cyberspace will be another great power-game space besides land, sea, air, and outer spaces.

In China, a government report has clearly proposed that cyberspace, as well as deep sea and deep space, are key areas of the national core interests. The lag behind in the field of big data research and applications not only means the loss of its industrial strategic advantage, but also suggests loopholes in its national security cyberspace. In this sense, the Big Data Research and Development Initiative[1] [13], announced by the United States in March 2012, is not only a strategic plan that promotes the US to continuously lead in the high-tech fields, but also a plan to protect its national security and advance its socio-economic development.

In general, the Western countries, represented by the United States, are moving under their national agenda towards a modernization of their national strength through big data research and applications. It is anticipated that future economic and political competitions among countries will be based on exploiting the potential of big data, among other traditional aspects. In short, the research and applications of big data are of strategic importance and significance for improving the competitiveness of any country.

2.2. Significance to industrial upgrades

Big data is currently a common problem faced by many industries, and it brings grand challenges to these industries' digitization and informationization. Research on common problems of big data, especially on breakthroughs of core technologies, will enable industries to harness the complexity induced by data interconnection and to master uncertainties caused by redundancy and/or shortage of data. Everyone hopes to mine from big data demand-driven information, knowledge and even intelligence and ultimately taking full advantage of the big value of big data. This means that data is no longer a byproduct of the industrial sector, but has become a key nexus of all aspects. In this sense, the study of common problems and core technologies of big data will be the focus of the new generation of IT and its applications. It will not only be the new engine to sustain the high growth of the information industry, but also the new tool for industries to improve their competitiveness.

For example, in recent years, cloud computing has rapidly evolved from a vague concept in the beginning to a mature hot technology. Many big companies, including Google, Microsoft, Amazon, Facebook, Alibaba,[2] Baidu,[3] Tencent,[4] and other IT giants, are working on cloud computing technologies and cloud-based computing services. Big data and cloud computing is seen as two sides of a coin: big data is a killer application of cloud computing, whereas cloud computing provides the IT infrastructure to big data. The tightly coupled big data and cloud computing nexus are expected to change the ecosystem of Internet, and even affect the pattern of the entire information industry.

2.3. Significance to scientific research

Big data has caused the scientific community to re-examine its methodology of scientific research [14] and has triggered a revolution in scientific thinking and methods.

It is well-known that the earliest scientific research in human history was based on experiments. Later on, theoretical science emerged, which was characterized by the study of various laws and theorems. However, because theoretical analysis is too complex and not feasible for solving practical problems, people began to seek simulation-based methods, which led to computational science.

The emergence of big data has spawned a new research paradigm; that is, with big data, researchers may only need to find or mine from it the required information, knowledge and intelligence. They even do not need to directly access the objects to be studied. In 2007, the late Turing Award winner, Jim Gray, depicted in his last speech the fourth paradigm of data-intensive scientific research [14], which separates data-intensive science from computational science. Gray believed that the fourth paradigm may be the only systemic way for solving some of the toughest global challenges we face today. In essence, the fourth paradigm is not only a change in the way of scientific research, but also a change in the way that people think [1].

2.4. Significance to emerging interdisciplinary research

Big data technologies and the corresponding fundamental research have become a research focus in academia. An emerging

[1] http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf.

[2] http://www.alibaba.com/.

[3] http://www.baidu.com/.

[4] http://www.tencent.com/.