

Chapter 10

Big Data and Business Decision Making

Marta Vidal

Complutense University of Madrid, Spain

Javier Vidal-García

University of Valladolid, Spain

Rafael Hernandez Barros

Complutense University of Madrid, Spain

ABSTRACT

The generalization of the big data and new techniques associated with the processing and analysis of large databases is revolutionizing both the scientific work as the management of companies. Applications such as personalized recommendations for Amazon have been a very significant improvement in the purchase experience for consumers. In this work we analyze the possibilities of big data to improve the services offered by companies and the customer experience and increase the efficiency of these companies. The work also examines some aspects associated with the use of big data such as the issues of data privacy and compliance with the regulations on the use of the information.

1. INTRODUCTION

This work tries to give a vision necessarily partial, given speed of change and the extension of the topic, on the possibilities offered by the techniques of big data in the field of business decision making. Certainly big data is not a new field, banks, shops, etc. have long used large amounts of data, for example, for studies of segmentation of their customers. Scientists have been using massive amounts of data for long time. In fact the information that is generated in the social networks or the internet is not the main source of the enormous accumulation of data that occurs every day. Scientific applications such as radio telescopes, the DNA sequencers or particle accelerators generate much more information. For example the LHC (Large Hadron Collider) of CERN produces a petabyte of data every second from sensors that

DOI: 10.4018/978-1-5225-0182-4.ch010

Big Data and Business Decision Making

capture all the emissions produced by the collision of particles. From a technological point of view the problem is not the data but that the speed at which we can store the information as it does not progress at the same pace as the information that is generated.

The contributions of the internet (search tool and its application to translators, predictions etc.) and expansion of social networks has had a great impact on what we might call the big data social (analysis of feelings, textual analysis, etc.). But in fact, most of the data that is generated on the planet is produced by the interaction of machines (for example sensors) with other machines. At the moment there is great excitement about the almost unlimited possibilities of big data to improve the productivity and the welfare of the people. In this perspective it is very frequent that in business meetings appear allusions to big data. But the description of the current situation has been characterized by an ironic form by Dan Ariely: big data is like adolescent sex, all talk about it, nobody really knows how to do so, everyone thinks that all the others are doing and, therefore, all over the world ensures that they also do.

2. BACKGROUND

At present many lines of research in computational sciences and many new businesses refers to the use of big data. The definition of big data is complex because many of the techniques that accompany these massive databases were known by other names such as data mining. By big data normally means the construction, organization and use of enormous amounts of data to extract relationships or create new forms of value in markets, organizations, public services, etc. This definition should be qualified to understand better the importance of the big data since the domain of these techniques is not only related to the size of the databases that use. In the first place when you think of big data is thought to have an enormous volume of information.

Secondly an application of big data involves the aggregation of information from various sources, which make it particularly important to the process of management and merging of data. The data can come from sensors, GPS of millions of phones, clicks, logs of servers, emails, etc. Therefore it is not a question of numeric data arranged in a standard fashion (for example in tables). The data is very heterogeneous and can include images, texts, sounds, etc. Companies also, by the own heterogeneity of the data, usually prevent data stored in fixed structures such as the classical relational database. The management of the data is done using NoSQL systems (not only SQL) as opposed to the traditional language of SQL queries. This tool is essential when companies work with many gigabytes of data or millions of observations with heterogeneous formats and whose structure may change in time and need to be easily scalable. Some of the tools used to manipulate big data are becoming the industry standard as Hadoop, MapReduce, Pig, None, OpenRefine, Hive, HBase, Mahout, ZooKeeper or Impala. The vast majority of these tools have as objective to allow the parallel processing necessary when working with huge databases.

Thirdly the information used usually has a very heterogeneous level of signal-noise ratio, although there is much more noise than in typical applications that use administrative data, surveys or internal information systems of business organizations (Grossman & Siegel (2014)). Fourthly the objective of the techniques of big data in general is not to discover causalities but produce predictive models. By contrast with the fundamental vision that is explained in the courses of traditional statistics and econometrics, in big data only matters correlations while the causality is irrelevant. Finally, the use and analysis of information tends to occur at a very high speed (Kruschwitz (2011)).

16 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/big-data-and-business-decision-making/157690

Related Content

Fishing Vessel Type Recognition Based on Semantic Feature Vector

Junfeng Yuan, Qianqian Zhang, Jilin Zhang, Youhuizi Li, Zhen Liu, Meiting Xue and Yan Zeng (2024). *International Journal of Data Warehousing and Mining* (pp. 1-18).

www.irma-international.org/article/fishing-vessel-type-recognition-based-on-semantic-feature-vector/349222

Critical and Future Trends in Data Mining: A Review of Key Data Mining Technologies/Applications

Jeffrey Hsu (2003). *Data Mining: Opportunities and Challenges* (pp. 437-452).

www.irma-international.org/chapter/critical-future-trends-data-mining/7613

The Concept of Big Data in Bureaucratic Service Using Sentiment Analysis

Lia Muliawaty, Kamal Alamsyah, Ummu Salamah and Dian Sa'adillah Maylawati (2022). *Research Anthology on Implementing Sentiment Analysis Across Multiple Disciplines* (pp. 1189-1202).

www.irma-international.org/chapter/the-concept-of-big-data-in-bureaucratic-service-using-sentiment-analysis/308539

A Comparative Study on Medical Diagnosis Using Predictive Data Mining: A Case Study

Seyed Jaleleddin Mousavirad and Hossein Ebrahimpour-Komleh (2014). *Data Mining and Analysis in the Engineering Field* (pp. 327-360).

www.irma-international.org/chapter/a-comparative-study-on-medical-diagnosis-using-predictive-data-mining/109989

An Efficient Code-Embedding-Based Vulnerability Detection Model for Ethereum Smart Contracts

Zhigang Xu, Xingxing Chen, Xinhua Dong, Hongmu Han, Zhongzhen Yan, Kangze Ye, Chaojun Li, Zhiqiang Zheng, Haitao Wang and Jiaxi Zhang (2023). *International Journal of Data Warehousing and Mining* (pp. 1-23).

www.irma-international.org/article/an-efficient-code-embedding-based-vulnerability-detection-model-for-ethereum-smart-contracts/320473