Preface

Big Data and Management Science has been designed to synthesize the analytic principles with business practice and Big Data. Specifically, the book provides an interface between the main disciplines of engineering/technology and the organizational, administrative, and planning abilities of management. It is complementary to other sub-disciplines such as economics, finance, marketing, decision and risk analysis.

This book is intended for engineers, economists, and researchers who wish to develop new skills in management or for those who employ the management discipline as part of their work. The authors of this volume describe their original work in the area or provide material for case studies that successfully apply the management discipline in real-life situations where Big Data is also employed.

The recent advances in handling large data have led to increasingly more data being available, leading to the advent of Big Data. The volume of Big Data runs into petabytes of information, offering the promise of valuable insight. Visualization is the key to unlocking these insights; however, repeating analytical behaviors reserved for smaller data sets runs the risk of ignoring latent relationships in the data, which is at odds with the motivation for collecting Big Data. Chapter "Visualizing Big Data: Everything Old Is New Again" focuses on commonly used tools (SAS, R, and Python) in aid of Big Data visualization to drive the formulation of meaningful research questions. It presents a case study of the public scanner database Dominick's Finer Foods, containing approximately 98 million observations. Using graph semiotics, it focuses on visualization for decision making and explorative analyses. It then demonstrates how to use these visualizations to formulate elementary-, intermediate-, and overall-level analytical questions from the database.

The development of Big Data applications is closely linked to the availability of scalable and cost-effective computing capacities for storing and processing data in a distributed and parallel fashion, respectively. Cloud providers already offer a portfolio of various cloud services for supporting Big Data applications. Large companies such as Netflix and Spotify already use those cloud services to operate their Big Data applications. Chapter "Managing Cloud-Based Big Data Platforms: A Reference Architecture and Cost Perspective" proposes a generic reference to architecture that implements Big Data applications based on state-of-the-art cloud services. The applicability and implementation of our reference to architecture is demonstrated for three leading cloud providers. Given these implementations, we analyze how main pricing schemes and cost factors can be used to compare respective cloud services. This information is based on a Big Data streaming use case. Derived findings are essential for cloud-based Big Data management from a cost perspective.

Most of the information about Big Data has focused on the technical side of the phenomenon. Chapter "The Strategic Business Value of Big Data" makes the case that business implications of utilizing Big Data are crucial to obtain a competitive advantage. To achieve such objective, the organizational impacts of Big Data for today's business competition and innovation are analyzed in order to identify different strategies a company may implement, as well as the potential value that Big Data can provide for organizations in different sectors of the economy and different areas inside such organizations. In the same vein, different Big Data strategies a company may implement toward its development are stated and suggestions regarding how enterprises such as businesses, nonprofits, and governments can use data to gain insights and make more informed decisions. Current and potential applications of Big Data are presented for different private and public sectors, as well as the ability to use data effectively to drive rapid, precise and profitable decisions.

Chapter "A Review on Big Data Security and Privacy in Healthcare Applications" considers the term Big Data and its usage in healthcare applications. With the increasing use of technologically advanced equipment in medical, biomedical, and healthcare fields, the collection of patients' data from various hospitals is also becoming necessary. The availability of data at the central location is suitable so that it can be used in need of any pharmaceutical feedback, equipment's reporting, analysis and results of any disease, and many other uses. Collected data can also be used for manipulating or predicting any upcoming health crises due to any disaster, virus, or climate change. Collection of data from various health-related entities or from any patient raises serious questions upon leakage, integrity, security, and privacy of data. The questions and issues are highlighted and discussed in the last section of this chapter to emphasize the broad pre-deployment issues. Available platforms and solutions are also discussed to overcome the arising situation and question the prudence of usage and deployment of Big Data in healthcare-related fields and applications. The available data privacy, data security, users' accessing mechanisms, authentication procedures, and privileges are also described.

Chapter "What Is Big Data" consists of three parts. The first section describes what Big Data is, the concepts of Big Data, and how Big Data arose. Big Data affects scientific schemes. It considers the limitations of predictions by using Big Data and a relation between Big Data and hypotheses. A case study considers an electric power of Big Data systems. The next section describes the necessity of Big Data. This is a view that applies aspects of macroeconomics. In service science capitalism, measurements of values of products need Big Data. Service products are classified into stock, flow, and rate of flow change. Immediacy of Big Data implements and makes sense of each classification. Big Data provides a macroeconomic model with behavioral principles of economic agents. The principles have mathematical representation with high affinity of correlation deduced from Big Data. In the last section, we present an explanation of macroeconomic phenomena in Japan since 1980 as an example of use of the macroeconomic model.

Chapter “Big Data for Conversational Interfaces: Current Opportunities and Prospects” is on conversational technologies. As conversational technologies develop, more demands are placed upon computer-automated telephone responses. For instance, we want our conversational assistants to be able to solve our queries in multiple domains, to manage information from different usually unstructured sources, to be able to perform a variety of tasks, and understand open conversational language. However, developing the resources necessary to develop systems with such capabilities demands much time and effort. For each domain, task, or language, data must be collected and annotated following a schema that is usually not portable. The models must be trained over the annotated data, and their accuracy must be evaluated. In recent years, there has been a growing interest in investigating alternatives to manual effort that allow exploiting automatically the huge amount of resources available in the Web. This chapter describes the main initiatives to extract, process, and contextualize information from these Big Data rich and heterogeneous sources for the various tasks involved in dialog systems, including speech processing, natural language understanding, and dialog management.

In Chapter “Big Data Analytics in Telemedicine : A Role of Medical Image Compression,” Big Data analytics which is one of most rapidly expanding fields has started to play a vital role in the field of health care. A major goal of telemedicine is to eliminate unnecessary traveling of patients and their escorts. Data acquisition, data storage, data display and processing, and data transfer represent the basis of telemedicine. Telemedicine hinges on transfer of text, reports, voice, images, and video between geographically separated locations. Out of these, the simplest and easiest is through text, as it is quick and simple to use, since sending text requires very little bandwidth. The problem with images and videos is that they require a large amount of bandwidth for transmission and reception. Therefore, there is a need to reduce the size of the image that is to be sent or received, i.e., data compression is necessary. This chapter deals with employing prediction as a method for compression of biomedical images. The approach presented in this chapter offers great potential in compression of the medical image under consideration, without degrading the diagnostic ability of the image.

A Big Data network design with risk-averse signal control optimization (RISCO) is considered to regulate the risk associated with hazmat transportation and minimize total travel delay. A bi-level network design model is presented for RISCO subject to equilibrium flow. A weighted sum risk equilibrium model is proposed in Chapter "A Bundle-Like Algorithm for Big Data Network Design with Risk-Averse Signal Control Optimization" to determine generalized travel cost at lower level problem. Since the bi-objective signal control optimization is generally non-convex and non-smooth, a bundle-like efficient algorithm is presented to solve the equilibrium-based model effectively. A Big Data bounding strategy is developed in Chapter "A Bundle-Like Algorithm for Big Data Network Design with Risk-Averse Signal Control Optimization" to stabilize solutions of RISCO with modest computational efforts. In order to investigate the computational advantage of the proposed algorithm for Big Data network design with signal optimization, numerical comparisons using real data example and general networks are made with current best well-known algorithms. The results strongly indicate that the proposed algorithm becomes increasingly computationally comparative to best known alternatives as the size of network grows.

Chapter "Evaluation of Evacuation Corridors and Traffic Management Strategies for Short-Notice Evacuation" presents a simulation study of the large-scale traffic data under a short-notice emergency evacuation condition due to an assumed chlorine gas spill incident in a derailment accident in the Canadian National (CN) Railway's railroad yard in downtown Jackson, Mississippi by employing the dynamic traffic assignment simulation program DynusT. In the study, the effective evacuation corridor and traffic management strategies were identified in order to increase the number of cumulative vehicles evacuated out of the incident-affected protective action zone (PAZ) during the simulation duration. An iterative three-step study approach based on traffic control and traffic management considerations was undertaken to identify the best strategies in evacuation corridor selection, traffic management method, and evacuation demand staging to relieve heavy traffic congestions for such an evacuation.

Chapter "Analyzing Network Log Files Using Big Data Techniques" considers the service to 26 buildings with more than 1000 network devices (wireless and wired) and access to more than 10,000 devices (computers, tablets, smartphones, etc.) which generate approximately 200 MB/day of data that is stored mainly in the DHCP log, the Apache HTTP log, and the Wi-fi log files. Within this context, Chapter "Analyzing Network Log Files Using Big Data Techniques" addresses the design and development of an application that uses Big Data techniques to analyze those log files in order to track information on the device (date, time, MAC address, and georeferenced position), as well as the number and type of network accesses for each building. In the near future, this application will help the IT department to analyze all these logs in real time.

Finally, Chapter "Big Data and Earned Value Management in Airspace Industry" analyzes earned value management (EVM) for project management. Actual cost and earned value are the parameters used for monitoring projects. These parameters are compared with planned value to analyze the project status. EVM covers scope, cost, and time and unifies them in a common framework that allows evaluation of project health. Chapter "Big Data and Earned Value Management in Airspace Industry" aims to integrate the project management and the Big Data. It proposes an EVM approach, developed from a real case study in aerospace industry, to simultaneously manage large numbers of projects.

Ciudad Real, Spain — Fausto Pedro García Márquez
Philadelphia, PA, USA — Benjamin Lev

http://www.springer.com/978-3-319-45497-9

Big Data Management
García Márquez, F.P.; Lev, B. (Eds.)
2017, XVI, 267 p. 107 illus., 38 illus. in color., Hardcover
ISBN: 978-3-319-45497-9