

Chapter 18

Descriptive and Predictive Analytical Methods for Big Data

Sema A. Kalaian

Eastern Michigan University, USA

Rafa M. Kasim

Indiana Tech University, USA

Nabeel R. Kasim

University of Michigan, USA

ABSTRACT

Data analytics and modeling are powerful analytical tools for knowledge discovery through examining and capturing the complex and hidden relationships and patterns among the quantitative variables in the existing massive structured Big Data in efforts to predict future enterprise performance. The main purpose of this chapter is to present a conceptual and practical overview of some of the basic and advanced analytical tools for analyzing structured Big Data. The chapter covers descriptive and predictive analytical methods. Descriptive analytical tools such as mean, median, mode, variance, standard deviation, and data visualization methods (e.g., histograms, line charts) are covered. Predictive analytical tools for analyzing Big Data such as correlation, simple- and multiple- linear regression are also covered in the chapter.

INTRODUCTION

Data analytics tools are used in variety of disciplines and fields of study such as business, engineering, information technology, environmental studies, information systems, health informatics, and other disciplines and fields of study. The demand for effective and sophisticated knowledge discovery using data analytical tools (descriptive, predictive, and perspective analytics) have been grown exponentially over the years as a result of the rise of using the web and mobile communication devices (e.g., mobile phones, iPads, GPS) to collect massive amount of data as well as technological advances in computer processing power including data storage, data warehouses, and integrated systems capabilities. Such

DOI: 10.4018/978-1-5225-7501-6.ch018

data is most often referred to as “Big Data” and characterized with the following four main features: Volume; Variety; Velocity; and Value.

Discovering knowledge using data analytics tools help business executives and leaders of for-profit and nonprofit organizations make informed data-based decisions to solve complex organizational and enterprise problems. For example, the survival of businesses and organizations in a knowledge-and-data driven economy is derived from the ability to transform large quantities of data and information to knowledge (Kalaian & Kasim, 2015). However, a decade ago, most such data was either not collected or entirely overlooked as a key resource for enterprise success because lack of knowledge and understanding of the value of such information and knowledge (Hair, 2007) and lack of computer storage, processing, and computing capabilities to handle such massive amount of data. However, the ability to design, develop, analyze, and implement a Big Data analytical application is directly dependent on the technical knowledge about of the architecture of the storage, processing, networking, and computing platforms from both of a hardware and software perspectives (Loshin, 2013).

However, most enterprises and businesses need to transform the massive amount of the collected data into intelligent information (knowledge) and insights about the characteristics and the underlying structure of the data such as trends, patterns, and relationships. Consequently, the intelligent information can be used to create a holistic and comprehensive view of an enterprise to make smart and informed data-based competitive enterprise decisions and strategic planning, strategic enterprise performance improvements, data-based and analytics-based competitive actions in delivering performance gains, and predictions of future organizational performance to gain competitive and global advantage.

However, conducting Big Data analytics depends on many factors and one of these key factors is the data scientist’s ability to analyze the massive collected data using the most appropriate Big Data analytical tools to make valid future enterprise performance decisions and predictions. The collected Big Data could be structured (e.g., numerical data such as total sales, transaction amount, transaction time of the day, etc.) or unstructured (e.g., texts, videos, images, photos, audio recordings, etc.). In other words, a data scientist or analyst must choose the proper descriptive and predictive analytical tools to analyze the data at hand to draw valid conclusions about the characteristics of the “Big Data” and then make valid future predictions of, for example, enterprise performance. The focus of this chapter is on descriptive and predictive analytical methods for analyzing structured data. Accordingly, the chapter will be organized into two major sections. In section I, descriptive analytical tools are covered. In section II, predictive analytical tools are covered.

DESCRIPTIVE ANALYTICS

Descriptive analytics are statistical methods in the data analytics including Big Data analytics toolbox for describing, summarizing, and visualizing massive amount of data that can be used for data discovery purposes. Any collected data is usually overwhelming and uninformative and it needs to be cleaned, organized, and summarized. In situations when we have massive amount of data, it is necessary to use descriptive analytics to summarize the data to gain insights about the characteristics of the data. The descriptive analytical methods help the data analyst to describe a data set by organizing, summarizing, and visualizing the information in the quantitative data using simple descriptive summary measures and visualization methods.

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/descriptive-and-predictive-analytical-methods-for-big-data/217837

Related Content

Dynamic Workflow Composition: Using Markov Decision Processes

Prashant Doshi, Richard Goodwin, Rama Akkiraju and Kunal Verma (2005). *International Journal of Web Services Research* (pp. 1-17).

www.irma-international.org/article/dynamic-workflow-composition/3052

Technical Architecture of Enabling Body of Knowledge System for Effective Learning and Information Dissemination

Liang-Jie Zhang and Jia Zhang (2013). *International Journal of Web Services Research* (pp. 41-62).

www.irma-international.org/article/technical-architecture-of-enabling-body-of-knowledge-system-for-effective-learning-and-information-dissemination/90265

Efficient Transport Bindings for Web Service Messages

Christian Werner, Carsten Buschmann, Tobias Jäcker and Stefan Fischer (2008). *Web Services Research and Practices* (pp. 83-109).

www.irma-international.org/chapter/efficient-transport-bindings-web-service/31211

Ontology and Service Oriented Programming

Bing Li and Wei-Tek Tsai (2005). *International Journal of Web Services Research* (pp. 33-68).

www.irma-international.org/article/ontology-service-oriented-programming/3063

Individual Doctor Recommendation in Large Networks by Constrained Optimization

Jibing Gong, Hong Cheng and Lili Wang (2015). *International Journal of Web Services Research* (pp. 16-28).

www.irma-international.org/article/individual-doctor-recommendation-in-large-networks-by-constrained-optimization/139755