

# Preface

Modern data science is related to massive data sets (BigData), machine learning, and cloud computing. There are multiple ways of understanding data science: (1) BigData with small cloud computational power, which requires very fast algorithms; (2) relatively small data sets with large cloud computational power, which we can compute by distributing data to the cloud without a very efficient algorithm; (3) BigData and with big cloud, which requires both techniques of algorithms and architectural infrastructures (new computing models); and (4) small data sets with small cloud, which just requires the standard methods.

This book contains state-of-the-art knowledge for researchers in data science. It also presents various problems in BigData and data science. We first introduce important statistical and computational methods for data analysis. For example, we discuss the principal component analysis for the dimension reduction of massive data sets. Then, we introduce graph theoretical methods such as GraphCut, the Laplacian matrix, and Google PageRank for data search and classification. We also discuss efficient algorithms, the hardness of problems involving various types of BigData, and geometric data structures. This book is particularly interested in the discussion of incomplete data sets and partial connectedness among data points or data sets. The second part of the book focuses on special topics, which cover topological analysis and machine learning, business and financial data recovery, and massive data classification and predication for high-dimensional data sets. Another purpose of this book is to challenge the major ongoing and unsolved problems in data science and provide some prospective solutions to these problems.

This book is a concise and quick introduction to the hottest topic in mathematics, computer science, and information technology today: data science. Data science first emerged in mathematics and computer science out of the research need for the numerous applications of BigData in the information technology, business, and medical industries. This book has two main objectives. The first objective of this book is to cover necessary knowledge in statistics, graph theory, algorithms, and computational science. There is also specific focus on the internal connectivity of incomplete data sets, which could be one of the central topics of future data science, unlike the existing method of data processing where data modeling is at the center.

The second focus of this book discusses major ongoing and unsolved problems in data science and provides some prospective solutions for these problems.

The book also collects some research papers from the talks given at the International Congress of Mathematics (ICM) 2014 Satellite Conference on Mathematical Foundation of Modern Data Sciences Computing, Logic, and Education, Dalian Maritime University, Dalian, China, which took place from July 27 to August 1, 2014. We are grateful to the Seoul ICM 2014 organization committee and National Science Foundation of China for their support. Many thanks go to Professor Reinhard Klette at the University of Auckland and Professor Wen Gao at Beijing University for giving excellent invited talks. Special thanks to Professors Shi-Qiang Wang (Beijing Normal University), Steven G. Krantz (Washington University), Shmuel Weinberger (University of Chicago), and Hanan Samet (University of Maryland) for their support. Special thanks also go to Dalian University of Technology, Dalian Maritime University, Southeast University of China, and University of the District of Columbia for their support to this conference.

This book has three parts. The first part contains the basics in data science; the second part mainly deals with computing, learning, and problems in data science; the third part is selected topics. Chapter 1: Introduction (L. Chen); Chap. 2: Overview of Basic Methods for Data Science (L. Chen); Chap. 3: Relationship and Connectivity of Incomplete Data Collection (L. Chen); Chap. 4: Machine Learning for Data Science (L. Chen); Chap. 5: Images, Videos, and BigData (L. Chen); Chap. 6: Topological Data Analysis (L. Chen); Chap. 7: Monte Carlo Methods and Their Applications in Big Data Analysis (H. Ji and Y. Li); Chap. 8: Feature Extraction via Vector Bundle Learning (R. Liu and Z. Su); Chap. 9: Curve Interpolation and Positivity-Preserving Financial Curve Construction (P. Huang, H. Wang, P. Wu, and Y. Li); Chap. 10: Advanced Methods in Variational Learning (J. Spencer and K. Chen); Chap. 11: On-line Strategies of Groups Evacuation from a Convex Region in the Plane (B. Jiang, Y. Liu, and H. Zhang); and Chap. 12: A New Computational Model of Bigdata (B. Zhu).

Washington, DC, USA  
Dalian, China

Li M. Chen  
Zhixun Su  
Bo Jiang

<http://www.springer.com/978-3-319-25125-7>

Mathematical Problems in Data Science

Theoretical and Practical Methods

Chen, L.M.; Su, Z.; Jiang, B.

2015, XV, 213 p. 64 illus., 42 illus. in color., Hardcover

ISBN: 978-3-319-25125-7