Preface

Human history has shown that the spread of civilization and expansion of economies can be largely attributed to transportation systems that connect countries, regions, cities, and neighborhoods. From horse-drawn carriages, to vehicles with internal combustion engines, to electric vehicles, and to future connected and automated vehicles, transportation is a rapidly advancing story making our lives and society more enriched and connected. Intelligent transportation systems (ITS) promise to make great strides in making our cities and regions smart and connected with other infrastructures such as the energy grid. ITS is becoming a part of Internet of things with new sensing, control, edge, and cloud computing technologies ready to be a part of smart cities and regions.

Transportation systems will continue to play a strategic role in our worldwide economy by delivering goods and people through increasingly complex, interconnected, and multimodal transportation systems. However, the complexities of modern transportation cannot be managed using yesterday's strategies and tools. ITS are characterized by increasingly complex data in heterogeneous formats, large volumes, nuances in spatial and temporal processes, and frequent real-time processing requirements. In addition, ITS will be enhanced with data collected from personal devices, social media, and services. Simple data processing, integration, and analytics tools do not meet the needs of complex ITS data processing tasks. The application of emerging data analytic systems and methods, with effective data collection and information distribution systems, provides opportunities that are required for building the ITS of today and tomorrow. Given the need for a new generation of professionals to work in data-intensive ITS, there is a need for a textbook that combines the diverse ITS-related data analytics topics. This book aims to prepare a skilled workforce, focusing on transportation engineering students and existing professionals, and also including data science students and professionals who will lead the planning, development, and maintenance of future ITS.

This book consists of 12 chapters covering diverse data analytics topics. Chapter 1 presents an overview of ITS and the data-intensive nature of diverse ITS applications. A summary of the sources and characteristics of ITS data including the relevance of ITS to data analytics is provided. In addition, a review of the US National ITS architecture is given as an example framework for ITS planning, design, and deployment, with an emphasis on the data analytics. An overview of ITS applications is provided to demonstrate the role of different stakeholders in ITS application deployment. This chapter ends with a brief history of ITS deployments around the world including emerging trends fueled by technological innovations such as automated vehicles.

Chapter 2 introduces data analytics fundamentals and their context in ITS. Descriptive, diagnostic, predictive, and prescriptive aspects of data analytics are described. Then, the evolution of data analytics solutions such as SQL analytics, visual analytics, big data analytics, and cognitive analytics are presented. Available open source data analytics tools and resources are also listed. This chapter concludes with a discussion about the future directions of data analytics for ITS.

Chapter 3 describes basic data science toolsets and sets the stage for the analytical techniques in the remainder of the book. Specific topics covered in this chapter are: (1) introduction to a basic statistical programming environments for complex data analytics, *R*, (2) overview of the *Research Data Exchange* ITS data repository, (3) fundamental concepts about structuring data in *R*, (4) techniques and libraries to ingest data files from external formats into *R*, (5) techniques and libraries to

extract data from online sources into *R*, and (6) a brief introduction to Big Data processing techniques.

Chapter 4 focuses on the data life cycle that enables researchers and practitioners to efficiently maintain data for real-time to long-term use. Data objects can be a collection of files and links or a database. The data life cycle encompasses a set of stages depending on the types of data. Moreover, there are different views on what are the stages of a data life cycle. This chapter aims to give an understanding of the life cycle of data.

Chapter 5 explores data infrastructure development solutions considering diverse ITS applications, their data workload characteristics, and corresponding requirements. An overview of infrastructures to support the requirements of data infrastructure capable of storing, processing, and distributing large volumes of data using different abstractions and runtime systems are presented. ITS application requirements are then mapped to a technical architecture for a data infrastructure. Different high-level infrastructures focusing on the different programming systems, abstraction, and infrastructures, and low-level infrastructure focusing on the storage and compute management are summarized.

Chapter 6 surveys ITS security and privacy issues. An overview of communications networks and the innovative applications in ITS are presented. Stakeholders within the automotive ecosystem and the assets they need to protect are identified. An attack taxonomy that describes attacks on ITS including connected vehicles is discussed. Existing attacks on connected vehicles are reviewed and mapped using the attack taxonomy. Finally, a discussion on existing and potential security and privacy solutions are presented.

Chapter 7 presents application of interactive data visualization concepts and tools integrated with data mining algorithms in the context of ITS. In the ITS domain, such systems are necessary to support decision making in large and complex data streams that are produced and consumed by different ITS infrastructures and components, such as traffic cameras, vehicles, and traffic management centers. An introduction to several key topics related to the design of data visualization systems for ITS is provided in this chapter. In addition, practical visualization design principles are discussed. This chapter concludes with a detailed case study involving the design of a multivariate visualization tool.

Chapter 8 discusses the application of system engineering principles in ITS. System engineering is used to allocate responsibilities, in the form of requirements, to both hardware and software on all platforms that participate in the ITS applications. A survey on the information needed as background for the data analysis-focused ITS systems development scenario is presented. In the development scenario, data communication requirements are identified and mapped those requirements using an Architecture Description Language (ADL). The ADL supports verification and analysis activities of the modeled system as discussed in chapter 8.

Chapter 9 focuses specifically on highway traffic safety data analysis. Different traffic safety analysis methods such as crash count/frequency modeling, safety effectiveness evaluation, economic appraisal, hot spot analysis, and injury severity modeling are explored. An overview of existing highway traffic safety research is provided first. Various methodologies that were used in these studies are summarized. Details of available data for highway traffic safety applications, including their limitations, are discussed. In addition, potential new data sources enabled by emerging trends such as connected and autonomous vehicles are explored.

Chapter 10 discusses the commonly used descriptive and predictive data analytics techniques in ITS applications in the context of intermodal freight transportation. These techniques cover the full

spectrum of univariate, bivariate, and multivariate analyses. This chapter also demonstrates how to apply these techniques using the *R* statistical software package.

Chapter 11 provides an overview of the application of social media data in ITS applications. As social media platforms such as Twitter, INSTAGRAM, and Facebook include postings about people's daily activities, including travel, they have become a rich source of data for supporting transportation planning and operations. Specific topics explored in this chapter are: (1) social media data characteristics, (2) a review of the most recent social media data analysis tools and algorithms, (3) a brief overview of the emerging social media applications in transportation, and (4) future research challenges and potential solutions.

Chapter 12 presents basic concepts of the machine learning methods and their application in ITS applications. This chapter discusses how machine learning methods can be utilized to improve performance of transportation data analytics tools. Selected machine learning methods, and importance of quality and quantity of available data are discussed. A brief overview of selected data preprocessing and machine learning methods for ITS applications is provided. An example is used to illustrate the importance of using machine learning method in data-driven transportation system.

This book presents data analytics fundamentals for ITS professionals, and highlights the importance of data analytics for planning, operating, and managing of future transportation systems. The data analytics areas presented in this book are useful for stakeholders involved in ITS planning, operation, and maintenance. The chapters are sufficiently detailed to communicate the key aspects of data analytics to transportation professionals anywhere in the workforce, whether in developed or developing countries.

This book can serve as a primary or supplemental textbook for upper-level undergraduate and graduate course on data analytics for ITS and can be adopted for analytics courses in many engineering disciplines, such as civil engineering, automotive engineering, computer science, and electrical and computer engineering. This book also presents the fundamentals of data analytics for ITS in a high-level, yet practice-oriented approach. The style of presentation will help ITS and related professionals around the world to use this book as a reference. The motivation of the editors for presenting this book is to inspire transportation system innovations that will enhance safety, mobility, and environmental sustainability with the use of data analytics as an important tool in the ITS cyber-physical domain.

Mashrur Chowdhury, Amy Apon and Kakan Dey