Chapter 1

Big Data and Transport Analytics: An Introduction

Constantinos Antoniou*, Loukas Dimitriou† and Francisco Câmara Pereira‡
**Department of Civil, Geo and Environmental Engineering, Technical University of Munich, Munich, Germany, †Laboratory for Transport Engineering, Department of Civil and Environmental Engineering, University of Cyprus, Nicosia, Cyprus, ‡Department of Management Engineering, Technical University of Denmark (DTU), Lyngby, Denmark*

Chapter Outline

1 INTRODUCTION

The aim of this book is to contribute to the question of how the transportation profession and research community can benefit from the new era of Big Data and Data Science, the opportunities that arise, the new threats that are emerging and how old and new challenges can be addressed by the enormous quantities of information that are foreseen to be available.

The current era can be characterized by three main components:

1. an unprecedented availability of (structured and unstructured) information, collected through traditional sources/sensors, but also by the extensive wealth of nontraditional sources, like internet-of-things and crowdsourcing;
2. a vast expansion of computational means (hardware and—most significantly—paradigms) exceeding Moore's law (Moore, 1965); and
3. the development of new powerful computational methods able to treat the challenges of extensive information, able to be executed only by powerful computational means (interconnected and cloud integrated).

These three elements triggered a tremendous boost in inspiration and incentives for new developments for business and industrial applications, in the associated research community, as well as in social and governmental organizations overall. The stage has changed.

Mobility Patterns, Big Data and Transport Analytics. https://doi.org/10.1016/B978-0-12-812970-8.00001-4

This constitutes the new vibrant scientific area of Data Science, adding a new data-driven analytical paradigm that combines the existing traditional three, viz., the empirical, the theoretical, and the computational. As any newcomer, Data Science has been received by many with some reluctance (Pigliucci, 2009; Milne and Watling, 2017), but by others as a path to new (and easy) revelations.

Famously, Chris Anderson declared that this is the "End of Theory,"[1] following the long tradition of human ambition for conquering knowledge and future, starting from the biblical "Tree of Knowledge," to statements of prolific figures of science like (purportedly) Charles Holland Duell's "Everything that can be invented has been invented," Lord Kelvin's "There is nothing new to be discovered in physics now; All that remains is more and more precise measurement" and David Hilbert's "We must know; We will know!," until all of them to be defeated (e.g., Gödel, 1931).

However, Anderson's (2008) statement that "Correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all," naively taken, may lead to the misconception that the fundamental use of models (hypothesis testing and explanatory analysis) is obsolete. Of course, this is not the essence of data analysis, though data-centric analysis has an impact on the experimental design, the type of information that is used for a particular research purpose, and the type of confirmatory criteria used for evaluating results. These differences in data use (in data availability and volume) and model building (in models' typology and fundamental assumptions) signify a turning point in scientific reasoning, requiring new theoretical and practical developments for treating the new scientific threats, as well as a preparation of the new generation of scientists, able to appropriately handle the new "tool(s)" that will increasingly become available.

Focusing in the field of transportation systems analysis, Data Science endeavors suit well, in their characteristics, which–succinctly–include:

- *Complex and Large scale*, composed by multiple distinctive units, arranged in multiple sequences, layers or parallel operations;
- *Spatially distributed*, establishing connectivity and service among remote locations by a synthesis of supply means (transport infrastructure and transport modes);
- *Multiple-agents engagement*, involved in cooperative, noncooperative, and competitive relationships among them and the transport infrastructure;
- *Dynamic/Transient*, since transport is by definition a dynamic phenomenon of movement in space and time; and
- *Stochastic*, since the transport operations stand for the manifestation of the decision-making process of agents (travelers, shippers, carriers, etc.) with different characteristics, properties, opportunities, "flavors" and criteria,

1. Wired, 06.23.08.

while decisions are made in a fluctuating environment in terms of the physical, economic, and other elements.

The above fundamental characteristics of the transportation systems comprise sources of complexity, reflected as inaccuracies (or failures) of the typical/traditional analytical paradigms, especially when applied in real-world circumstances. The use of Big Data, treated within the new analytical field of Data Science, in our view stands for a promising new era for understanding and managing existing and future transportation phenomena.

The effective exploitation of the Big Data and Data Science "promises" depends on the rate of endorsement of emerging methods and applications by the relevant scientific and industrial community. It should be highlighted that the general public (such as the end-users and the markets) anticipate new developments, with the community of "*early-adopters*" growing rapidly.

But what are the characteristics of the relevant contemporary (and future) transportation scientist? How should the new generation of transportation scientists be equipped? Is it all about data handling/processing/analysis? The view that can be identified throughout this book reflects the idea that the strong scientific background on the field, topic, or system is a compulsory prerequisite for testing or adopting data-centric applications. This is far from the so-called Black-Box approaches and the jeopardies involved in such cases or applications. Advanced data analysis and Data Science concepts stand for an additional tool of the transportation professional, who should be formally prepared (possibly by dedicated programs) for embracing them. However, this should not be viewed as a shortcut to avoid the fundamentals.

Finally, the idea for this book was conceived during a Summer School that was organized by the Editors and held on these topics in June 2016 in the premises of the University of Cyprus, Lab for Transport Engineering, with the participation of most of the (co)authors. During the Summer School, the multidisciplinary combination of both the instructors, and the attendees, became immediately evident. This pluralism of ideas, approaches, and concepts is reflected here. We are confident that the readers will benefit from the contents of this book and will enjoy this guided trip through the different topics, models, and applications aiming to cover some of the most important fields of the transportation profession, where Big Data applications have matured enough.

2 BOOK STRUCTURE

The book is structured, instead of the form of a textbook, to provide a guided trip of the reader from introductory concepts about the use of Big Data and Data Science methods in transportation, to the presentation of indicative (though mature) applications. In this way, the reader may gain a helpful insight on how concepts may be treated with new methods, such as to develop innovative applications. We think that this innovation-oriented process may be more

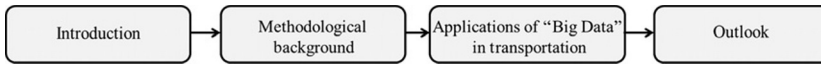

FIG. 1 Flowchart of the book structure.

inspiring and has timely importance rather than the presentation of a "strict" closed-form/self-contained textbook of specific applications on some topics of mobility. For achieving this, the book is divided into four Sections; the introductory, the methodological, the applications, and outlook (Fig. 1).

In the current introductory part (this chapter), the general view of the book on the use and value of data-centric analysis is provided. Then, at the second part of the book, a review and presentation of fundamentals on machine-learning methods is provided (Chapter 2), while a discussion about the combination of theory-driven and data-driven methods is offered in Chapter 3. Then, the stage of mobility analysis, human activities, and the living structure within the geographical space is offered in Chapter 4. Issues regarding data preparation (Chapter 5) and data visualization (Chapter 6) are providing important preparation on prospective transportation data-analytics professionals and researchers. The preparatory part of the book comprises the theoretical underpinnings on the integration of model-based machine-learning approaches in transportation (Chapter 7) and the use of nontraditional textual data for analyzing mobility (Chapter 8). After having read the methodological part, the reader will be equipped with essential methodological tools in order to be able to better follow the applications-oriented part. In detail, indicative applications of nontraditional examples using data-analytic approaches are selected, facilitating the understanding of how new methods have been adopted in order to analyze transportation demand and systems (Chapters 9, 10 and 15; Chapter 11), road safety (Chapter 12), mobility patterns (Chapter 13), and transport infrastructure (Chapter 14). By the end of the book's third part the reader will have gained knowledge on applications of currently state-of-the-art methods in various elements of the transportation field and hopefully inspirations for extending or improving the use of Big Data and Data Science methods in the field. The fourth and final part of book provides an outlook on the use of advanced data analytics methods in transport, aiming to offer a useful foresight to potential transport professionals, developers, and researchers.

SPECIAL ACKNOWLEDGMENTS

The Editors of this book would like to gratefully acknowledge the patient and helpful support of the Editorial Managers (especially Mr. Tom Stover, Acquisitions Editor, Transport and Ms. Naomi Robertson) throughout the development process, as they provided their valuable experience and organization facilities for putting together and concluding this work.

Furthermore, the Editors would like to express their sincere and deep gratitude to all distinguished authors that contributed to this effort and agreed to share their academic and professional experience with the scientific community through this volume. Without their invaluable contribution, this book would not be possible.

REFERENCES

C. Anderson, 2008. The End of Theory: The Data Deluge Makes the Scientific Method Obsolete.

Gödel, K., 1931. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, I. Monatsh. Math. Phys. 38, 173–198.

Milne, D., Watling, D., 2017. Big data and understanding change in the context of planning transport systems. J. Transp. Geogr. https://doi.org/10.1016/j.jtrangeo.2017.11.004.

Moore, G., 1965. Cramming more components onto integrated circuits. Electronics. 38 (8).

Pigliucci, M., 2009. The end of theory in science? EMBO Rep. 10 (6), 534. Science and Society, Opinions.

FURTHER READING

D. Hilbert, 1930. Address to the Society of German Scientists and Physicians, in Königsberg.