

Preface

The author has worked for 12 years as an information scientist in interdisciplinary fields involving physics, economics and information studies and has written works that specifically fall within econophysics and information sciences. The specific intention behind this book is to contribute to “econoinformatics”.

The data on human societies was partial and limited at the beginning of twenty-first century. However, current data availability has improved remarkably. As a result, researchers in various fields such as economics, finance, marketing, data mining, sociology, physics and information sciences have a similar interest in data on our society and study societal issues using a large amount of these data. They have opened up a new paradigm of studies on society that is described by a single keyword: data. In this book, these new emerging fields are termed “Applied Data-Centric Social Sciences”, which in association with data use, and this book has been written to share a vision of these fields.

Human society often shows interesting properties, such as non-stationarity, synchrony and spatiotemporal patterns. In order to capture these properties, individual behaviour, social relationships among individuals and man-made artificial systems need to be understood.

Since the nineteenth century, social scientists have conducted empirical investigations to understand the characteristics of collective versus individual behaviour and relationships between individuals within a society. Concurrently, data on socio-economic technological systems have accumulated in various fields of study.

Since research topics within applied data-centric social sciences are wide and deep, this book attempts to introduce some fundamental segments of these fields, including several mathematical expressions and some techniques to handle a vast amount of data and computer analysis. This work is also based on several example studies of data-oriented investigation in which advanced mathematics is used to analyse and model several specific problems.

The fundamental philosophy underlying this book is that both mathematical and physical expressions should be used to express actual, real-world data with high accuracy and thereby understand data-generating mechanism.

In data-centric science method, thinking first starts from the data in a specific field. Next, a search is attempted for an adequate method or expression to investigate the data. Explanatory data analysis provides an improvement cycle

through data acquisition, data collection and data analysis to reach interpretation. This type of activity constructs a PDCA (plan-do-check-action) cycle, which is a part of data-centric science. Data preservation and data recycling may thus be examined over a longer time horizon, a process sometimes referred to as ‘data curation’.

Mathematics are useful tools to express processes and states of our socio-economic technological systems. If problems can be described by well-defined mathematical notations and modelled by expressions, then cyber-enabled techniques can be constructed—such as automated data collection, automatic data verification and optimisation techniques—by using both mathematical and physical expressions and models.

Data is defined as several numbers to describe physical quantities (e.g. length, weight, time and velocity) or a number of texts expressing the actual situations or processes. The data is collected to trace real-world situations (records) to transfer information on what is focused on in the real world to other people (communications). Why should the data be analysed? The dominant purpose of data analysis is deeply associated with decision-making. In general, humans want to know and understand processes and phenomena in more details when they have to make some decisions in actual environments where they live. Thus, the results of data analysis are used as information for decision-making in the real world.

Some contact points to actual society are present in data analysis on socio-economic technological systems. Legal issues on data-centric social sciences are also addressed, although some of these issues are still under discussion.

Furthermore, this book contains separate intentions behind each chapter. The first of these is how to describe mathematical and statistical methods for data analysis. The second is to look at the background of data-generating mechanism. The third is the motivation for applying data-centric sciences to socio-economic technological systems. The example studies using the data of a specific field are presented to help readers to understand the different situations of socio-economic technological systems and how mathematical and physical methods are applied to actual data. Activities in data-centric social sciences are addressed as much as possible.

To satisfy these intentions, this book is organised as follows. Chapter 1 discusses the concepts used to deal with data on socio-economic technological systems and the reliance of the data-centric social sciences on the data for social activities, relationships and behaviour. Chapter 2 explains the research framework of applied data-centric social sciences. Why do scientists acquire, collect and analyse data on human society? How can the results of data analysis be utilised? This chapter also discusses methodologies within applied data-centric social sciences.

Chapter 3 introduces mathematical expressions used to describe societies, human behaviour and relationship. Several fundamental methods are explained, including statistical procedures, stochastic methods, network description and geographic information. Chapter 4 shows several methods of processing data with

computers. Database servers and parallel computation techniques are needed to handle large scale data.

Chapter 5 shows an example study of risk assessment in the foreign exchange market by using both q -Gaussian and Pearson type IV distributions, while Chap. 6 discusses a method to quantify states of the foreign exchange market by using a recursive segmentation procedure. Chapter 7 presents analysis of Japanese hotel booking data and quantifies the regional dependence of hotels, and Chap. 8 looks at relationships between flight ticket prices and their geodesic distance. Chapter 9 considers the relationship between electric power consumption per capita and economic performance (GDP per capita). Finally, Chap. 10 examines the future of the applied data-centric social sciences.

This book hopes, in this way, to encourage readers to acquire, collect, store, analyse and interpret data from socio-economic technological systems, in order to solve their own problems.

Kyoto, Zurich, 2013, 2014

Aki-Hiro Sato

Applied Data-Centric Social Sciences
Concepts, Data, Computation, and Theory

Sato, A.-H.

2014, XXIII, 281 p. 71 illus., 20 illus. in color., Hardcover

ISBN: 978-4-431-54973-4