Book Review
Data Mining Applications with R

Edited by Yanchang Zhao and Yonghua Cen, Elsevier Academic Press USA (2014)
ISBN:- 978-0-12-441511-8

Shabir A Bhat*

This book is collection of several articles written by authors who are professional in data mining and have conducted extensive research in their areas of specialization using R which is one of the mostly widely used data mining tool in business applications among several other open source software. It is not easy for beginners to find appropriate packages to use for their data mining tasks. It is difficult for experienced users to well in a way to find most optimum packages to solve their business problems and the best way to use them in data mining process of their applications. This book aims to facilitate using R in data mining applications by presenting real world applications in various domains. The book consist of 12 chapters and the review is given below:

The book starts with the presentation of an approach to analysis of large-scale time series sensor data collected from the electric power grid. This discussion is driven by our analysis of a real-world data set and, as such, does not provide a comprehensive exposition of either the tools used or the breadth of analysis appropriate for general time series data. Instead, this section provides the reader with sufficient information, motivation, and resources to address their own analysis challenges.It is reasonable to assume that there is a hidden process that explains the data we observe. Though we do not know the details of this process, we know that it is not completely random. Thus, mathematical models defined as parameters which can be used for this task are addressed.

As a specific case, the author investigates the culture of microblogging academics within the dynamics of a professional conference to gain insights into the key issues and debates emergent in this community and the transformative effects of using Twitter in academic contexts. Microblogging academics can be considered a type of online community which has its own rules, norms, and communicative behaviors. The author further stress on the unique and eventful characteristics of the 2011 meeting that make the related twitter content especially worth of investigation. These include organizational issues such as the session was convened in response to controversy surrounding the removal of the word "science" from the AAA's long-range plan statement in 2010.

The author stresses on the availability of open and structured data, digital libraries become an important source of data in recent data mining techniques. The inherent structure of data libraries become an important source of data comes with information about date, authorship, involved institutions, geographical context and large volumes of text.

A case study of the retailers has also been incorporated, where many retailers carry a myriad of products. It becomes an issue of matching the right product to the right person and is solved by Recommender Systems. The recommender lab package is extensively illustrated with practical examples. A number of different model builders are employed, looking at data mining in direct marketing. This theme of marketing and customer management is continued in the next chapter looking at the profiling of customers for insurance. A link to the data set used is provided in order to make it easy to follow along.

*Professor, Department of Management Studies, University of Kashmir

Chapter 8 critically discusses important task of feature selection in the context of identifying customers who may default on their bank loans. Various R packages are used and a selection of visualizations provide insights into the data. Travelers and their preferences for their hotels in the next chapter using Rfm tool which starts with the focus on some of the spatial and mapping capabilities of R for data mining. Spatial mapping and statistical analyses combine to provide insights into real estate pricing.

The book is rounded out with the application of data mining to the analysis of domain name system data. The aim is to deliver efficiencies for DNS servers. Cluster analysis using $k$means and $k$medoids forms the primary tool, and the authors again make effective use of very many different types of visualizations.

The authors of all the chapters of this book provide and share a breadth of insights, illustrated through the use of R. There is much to learn by watching masters at work, and that what we can gain from this book. The reader's focus should be on replicating the variety of analyses demonstrated throughout the book using their own data. There is so much pupil can learn about their own application from doing so.