Practical Graph Mining With R By Nagiza F Samatova

Practical Graph Mining with R
Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications
Graph Drawing and Network Visualization
A Practical Introduction
32nd International Conference, DEXA 2021, Virtual Event, September 27–30, 2021, Proceedings, Part I
Examples and Case Studies
Doing Meta-Analysis with R
7th Pacific-Asia Conference, PAKDD 2003. Seoul, Korea, April 30 - May 2, 2003, Proceedings
A Practical Introduction
R in Action
Behavior Analysis with Machine Learning Using R
Practical Skills with Python and Graphviz
22nd International Symposium, PADL 2020, New Orleans, LA, USA, January 20–21, 2020, Proceedings
Contributions Presented at the 16th UK Workshop on Computational Intelligence, September 7–9, 2016, Lancaster, UK
Machine Learning and Knowledge Discovery in Databases
Computational Statistics with R
Practical Aspects of Declarative Languages
Mining Complex Networks
Handbook of Statistics
Managing and Mining Graph Data
Unsupervised Machine Learning
Data analysis and graphics with R
European Conference, Antwerp, Belgium, September 15-19, 2008, Proceedings, Part I
Principles, Algorithms, and Applications
Inductive Logic Programming
Advancing Geographic Information Science: The Past and Next Twenty Years
Hands-On Graph Analytics with Neo4j
Mastering Social Media Mining with R
Collaborative Recommendations: Algorithms, Practical Challenges And Applications
Mining of Massive Datasets
Introduction to Data Mining
Text Mining with R
Deep Learning on Graphs
Graph Algorithms
Data Visualization
Laws, Tools, and Case Studies
Data Mining with R
A Hands-On Guide

R and Data Mining
Data Mining Applications with R

MORA SYLVIA

Practical Graph Mining with R Morgan & Claypool Publishers
This two-volume set, LNCS 12923 and 12924, constitutes the thoroughly refereed proceedings of the 5th International Conference on Database and Expert Systems Applications, DEXA 2021. Due to COVID-19 pandemic, the conference was held virtually. The 37 full papers presented together with 31 short papers in these volumes were carefully reviewed and selected from a total of 149 submissions. The papers are organized around the following topics: big data; data analysis and data modeling; data mining; databases and data management; information retrieval; prediction and decision support.
Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications Springer
Summary R in Action, Second Edition presents both the R language and the examples that make it so useful for business developers. Focusing on practical solutions, the book offers a crash course in statistics and covers elegant methods for dealing with messy and incomplete data that are difficult to analyze using traditional methods. You'll also master R's extensive graphical capabilities for exploring and presenting data visually. And this expanded second edition includes new chapters on time series analysis, cluster analysis, and classification methodologies, including decision trees, random forests, and support vector machines. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology Business pros and researchers thrive on data, and R speaks the language of data analysis. R is a powerful programming language for statistical computing. Unlike general-purpose tools, R provides thousands of modules for solving just about any data-crunching or presentation challenge you're likely to face. R runs on all important platforms and is used by thousands of major corporations and institutions worldwide. About the Book R in Action, Second Edition teaches you how to use the R language by presenting examples relevant to scientific, technical, and business developers. Focusing on practical solutions, the book offers a crash course in statistics, including elegant methods for dealing with messy and incomplete data. You'll also master R's extensive graphical capabilities for exploring and presenting data visually. And this expanded second edition includes new chapters on forecasting, data mining, and dynamic report writing. What's Inside Complete R language tutorial Using R to manage, analyze, and visualize data Techniques for debugging programs and creating packages OOP in R Over 160 graphs About the Author Dr. Rob Kabacoff is a seasoned researcher and teacher who specializes in data analysis. He also maintains the popular Quick-R website at statmethods.net. Table of Contents PART 1 GETTING STARTED Introduction to R Creating a dataset Getting started with graphs Basic data management Advanced data management PART 2 BASIC METHODS Basic graphs Basic statistics PART 3 INTERMEDIATE METHODS Regression Analysis of variance Power analysis Intermediate graphs Resampling statistics and bootstrapping PART 4 ADVANCED METHODS Generalized linear models Principal components and factor analysis Time series Cluster analysis Classification Advanced methods for missing data PART 5 EXPANDING YOUR SKILLS Advanced graphics with ggplot2 Advanced programming Creating a package Creating dynamic reports Advanced graphics with the lattice package available online only from manning.com/kabacoff2
Graph Drawing and Network Visualization John Wiley & Sons
Although there are several good books on unsupervised machine learning, we felt that many of them are too theoretical. This book provides practical guide to cluster analysis, elegant visualization and interpretation. It contains 5 parts. Part I provides a quick introduction to R and presents required R packages, as well as, data formats and dissimilarity measures for cluster analysis and visualization. Part II covers partitioning clustering methods, which subdivide the data sets into a set of k groups, where k is the number of groups pre-specified by the analyst. Partitioning clustering approaches include: K-means, K-Medoids (PAM) and CLARA algorithms. In Part III, we consider hierarchical clustering method, which is an alternative approach to partitioning clustering. The result of hierarchical clustering is a tree-based representation of the objects called dendrogram. In this part, we describe how to compute, visualize, interpret and compare dendrograms. Part IV describes clustering validation and evaluation strategies, which consists of measuring the goodness of clustering results. Among the chapters covered here, there are: Assessing clustering tendency, Determining the optimal number of clusters, Cluster validation statistics, Choosing the best clustering algorithms and Computing p-value for hierarchical clustering. Part V presents advanced clustering methods, including: Hierarchical k-means clustering, Fuzzy clustering, Model-based clustering and Density-based clustering.
A Practical Introduction World Scientific
R and Data Mining introduces researchers, post-graduate students, and analysts to data mining using R, a free software environment for statistical computing and graphics. The book provides practical methods for using R in applications from academia to industry to extract knowledge from vast amounts of data. Readers will find this book a valuable guide to the use of R in tasks such as classification and prediction, clustering, outlier detection, association rules, sequence analysis, text mining, social network analysis, sentiment analysis, and more. Data mining techniques are growing in popularity in a broad range of areas, from banking to insurance, retail, telecom, medicine, research, and government. This book focuses on the modeling phase of the data mining process, also addressing data exploration and model evaluation. With three in-depth case studies, a quick reference guide, bibliography, and links to a wealth of online resources, R and Data Mining is a valuable, practical guide to a powerful method of analysis. Presents an introduction into using R for data mining applications, covering most popular data mining techniques Provides code examples and data so that readers can easily learn the techniques Features case studies in real-world applications to help readers apply the techniques in their work
32nd International Conference, DEXA 2021, Virtual Event, September 27–30, 2021, Proceedings, Part I Morgan & Claypool Publishers

Data Mining: Concepts and Techniques provides the concepts and techniques in processing gathered data or information, which will be used in various applications. Specifically, it explains data mining and the tools used in discovering knowledge from the collected data. This book is referred as the knowledge discovery from data (KDD). It focuses on the feasibility, usefulness, effectiveness, and scalability of techniques of large data sets. After describing data mining, this edition explains the methods of knowing, preprocessing, processing, and warehousing data. It then presents information about data warehouses, online analytical processing (OLAP), and data cube technology. Then, the methods involved in mining frequent patterns, associations, and correlations for large data sets are described. The book details the methods for data classification and introduces the concepts and methods for data clustering. The remaining chapters discuss the outlier detection and the trends, applications, and research frontiers in data mining. This book is intended for Computer Science students, application developers, business professionals, and researchers who seek information on data mining. Presents dozens of algorithms and implementation examples, all in pseudo-code and suitable for use in real-world, large-scale data mining projects Addresses advanced topics such as mining object-relational databases, spatial databases, multimedia databases, time-series databases, text databases, the World Wide Web, and applications in several fields Provides a comprehensive, practical look at the concepts and techniques you need to get the most out of your data

Examples and Case Studies CRC Press

Discover Novel and Insightful Knowledge from Data Represented as a Graph Practical Graph Mining with R presents a "do-it-yourself" approach to extracting interesting patterns from graph data. It covers many basic and advanced techniques for the identification of anomalous or frequently recurring patterns in a graph, the discovery of groups or clusters of nodes that share common patterns of attributes and relationships, the extraction of patterns that distinguish one category of graphs from another, and the use of those patterns to predict the category of new graphs. Hands-On Application of Graph Data Mining Each chapter in the book focuses on a graph mining task, such as link analysis, cluster analysis, and classification. Through applications using real data sets, the book demonstrates how computational techniques can help solve real-world problems. The applications covered include network intrusion detection, tumor cell diagnostics, face recognition, predictive toxicology, mining metabolic and protein-protein interaction networks, and community detection in social networks. Develops Intuition through Easy-to-Follow Examples and Rigorous Mathematical Foundations Every algorithm and example is accompanied with R code. This allows readers to see how the algorithmic techniques correspond to the process of graph data analysis and to use the graph mining techniques in practice. The text also gives a rigorous, formal explanation of the underlying mathematics of each technique. Makes Graph Mining Accessible to Various Levels of Expertise Assuming no prior knowledge of mathematics or data mining, this self-contained book is accessible to students, researchers, and practitioners of graph data mining. It is suitable as a primary textbook for graph mining or as a supplement to a standard data mining course. It can also be used as a reference for researchers in computer, information, and computational science as well as a handy guide for data analytics practitioners.

Doing Meta-Analysis with R STHDA

Graphs naturally represent information ranging from links between web pages, to communication in email networks, to connections between neurons in our brains. These graphs often span billions of nodes and interactions between them. Within this deluge of interconnected data, how can we find the most important structures and summarize them? How can we efficiently visualize them? How can we detect anomalies that indicate critical events, such as an attack on a computer system, disease formation in the human brain, or the fall of a company? This book presents scalable, principled discovery algorithms that combine globality with locality to make sense of one or more graphs. In addition to fast algorithmic methodologies, we also contribute graph-theoretical ideas and models, and real-world applications in two main areas: •Individual Graph Mining: We show how to interpretably summarize a single graph by identifying its important graph structures. We complement summarization with inference, which leverages information about few entities (obtained via summarization or other methods) and the network structure to efficiently and effectively learn information about the unknown entities. •Collective Graph Mining: We extend the idea of individual-graph summarization to time-evolving graphs, and show how to scalably discover temporal patterns. Apart from summarization, we claim that graph similarity is often the underlying problem in a host of applications where multiple graphs occur (e.g., temporal anomaly detection, discovery of behavioral patterns), and we present principled, scalable algorithms for aligning networks and measuring their similarity. The methods that we present in this book leverage techniques from diverse areas, such as matrix algebra, graph theory, optimization, information theory, machine learning, finance, and social science, to solve real-world problems. We present applications of our exploration algorithms to massive datasets, including a Web graph of 6.6 billion edges, a Twitter graph of 1.8 billion edges, brain graphs with up to 90 million edges, collaboration, peer-to-peer networks, browser logs, all spanning millions of users and interactions.

7th Pacific-Asia Conference, PAKDD 2003. Seoul, Korea, April 30 - May 2, 2003, Proceedings Morgan & Claypool Publishers

Discover how graph algorithms can help you leverage the relationships within your data to develop more intelligent solutions and enhance your machine learning models. You'll learn how graph analytics are uniquely suited to unfold complex structures and reveal difficult-to-find patterns lurking in your data. Whether you are trying to build dynamic network models or forecast real-world behavior, this book illustrates how graph algorithms deliver value—from finding vulnerabilities and bottlenecks to detecting communities and improving machine learning predictions. This practical book walks you through hands-on examples of how to use graph algorithms in Apache Spark and Neo4j—two of the most common choices for graph analytics. Also included: sample code and tips for over 20 practical graph algorithms that cover optimal pathfinding, importance through centrality, and community detection. Learn how graph analytics vary from conventional statistical analysis Understand how classic graph algorithms work, and how they are applied Get guidance on which algorithms to use for different types of questions Explore algorithm examples with working code and sample datasets from Spark and Neo4j See how connected feature extraction can increase machine learning accuracy and precision Walk through creating an ML workflow for link prediction combining Neo4j and Spark

A Practical Introduction Springer

Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications brings

together all the information, tools and methods a professional will need to efficiently use text mining applications and statistical analysis. Winner of a 2012 PROSE Award in Computing and Information Sciences from the Association of American Publishers, this book presents a comprehensive how-to reference that shows the user how to conduct text mining and statistically analyze results. In addition to providing an in-depth examination of core text mining and link detection tools, methods and operations, the book examines advanced preprocessing techniques, knowledge representation considerations, and visualization approaches. Finally, the book explores current real-world, mission-critical applications of text mining and link detection using real world example tutorials in such varied fields as corporate, finance, business intelligence, genomics research, and counterterrorism activities. The world contains an unimaginably vast amount of digital information which is getting ever vaster ever more rapidly. This makes it possible to do many things that previously could not be done: spot business trends, prevent diseases, combat crime and so on. Managed well, the textual data can be used to unlock new sources of economic value, provide fresh insights into science and hold governments to account. As the Internet expands and our natural capacity to process the unstructured text that it contains diminishes, the value of text mining for information retrieval and search will increase dramatically. Extensive case studies, most in a tutorial format, allow the reader to 'click through' the example using a software program, thus learning to conduct text mining analyses in the most rapid manner of learning possible Numerous examples, tutorials, power points and datasets available via companion website on Elsevierdirect.com Glossary of text mining terms provided in the appendix

R in Action Chapman and Hall/CRC

Doing Meta-Analysis with R: A Hands-On Guide serves as an accessible introduction on how meta-analyses can be conducted in R. Essential steps for meta-analysis are covered, including calculation and pooling of outcome measures, forest plots, heterogeneity diagnostics, subgroup analyses, meta-regression, methods to control for publication bias, risk of bias assessments and plotting tools. Advanced but highly relevant topics such as network meta-analysis, multi-three-level meta-analyses, Bayesian meta-analysis approaches and SEM meta-analysis are also covered. A companion R package, dmetar, is introduced at the beginning of the guide. It contains data sets and several helper functions for the meta and metafor package used in the guide. The programming and statistical background covered in the book are kept at a non-expert level, making the book widely accessible. Features • Contains two introductory chapters on how to set up an R environment and do basic imports/manipulations of meta-analysis data, including exercises • Describes statistical concepts clearly and concisely before applying them in R • Includes step-by-step guidance through the coding required to perform meta-analyses, and a companion R package for the book

Behavior Analysis with Machine Learning Using R Simon and Schuster

R is open source statistical computing software. Since the R core group was formed in 1997, R has been extended by a very large number of packages with extensive documentation along with examples freely available on the internet. It offers a large number of statistical and numerical methods and graphical tools and visualization of extraordinarily high quality. R was recently ranked in 14th place by the Transparent Language Popularity Index and 6th as a scripting language, after PHP, Python, and Perl. The book is designed so that it can be used right away by novices while appealing to experienced users as well. Each article begins with a data example that can be downloaded directly from the R website. Data analysis questions are articulated following the presentation of the data. The necessary R commands are spelled out and executed and the output is presented and discussed. Other examples of data sets with a different flavor and different set of commands but following the theme of the article are presented as well. Each chapter predents a hands-on-experience. R has superb graphical outlays and the book brings out the essentials in this arena. The end user can benefit immensely by applying the graphics to enhance research findings. The core statistical methodologies such as regression, survival analysis, and discrete data are all covered. Addresses data examples that can be downloaded directly from the R website No other source is needed to gain practical experience Focus on the essentials in graphical outlays

Practical Skills with Python and Graphviz Springer

Data Mining Applications with R is a great resource for researchers and professionals to understand the wide use of R, a free software environment for statistical computing and graphics, in solving different problems in industry. R is widely used in leveraging data mining techniques across many different industries, including government, finance, insurance, medicine, scientific research and more. This book presents 15 different real-world case studies illustrating various techniques in rapidly growing areas. It is an ideal companion for data mining researchers in academia and industry looking for ways to turn this versatile software into a powerful analytic tool. R code, Data and color figures for the book are provided at the RDataMining.com website. Helps data miners to learn to use R in their specific area of work and see how R can apply in different industries Presents various case studies in real-world applications, which will help readers to apply the techniques in their work Provides code examples and sample data for readers to easily learn the techniques by running the code by themselves

22nd International Symposium, PADL 2020, New Orleans, LA, USA, January 20–21, 2020, Proceedings "O'Reilly Media, Inc."

Behavior Analysis with Machine Learning Using R introduces machine learning and deep learning concepts and algorithms applied to a diverse set of behavior analysis problems. It focuses on the practical aspects of solving such problems based on data collected from sensors or stored in electronic records. The included examples demonstrate how to perform common data analysis tasks such as: data exploration, visualization, preprocessing, data representation, model training and evaluation. All of this, using the R programming language and real-life behavioral data. Even though the examples focus on behavior analysis tasks, the covered underlying concepts and methods can be applied in any other domain. No prior knowledge in machine learning is assumed. Basic experience with R and basic knowledge in statistics and high school level mathematics are beneficial. Features: Build supervised machine learning models to predict indoor locations based on WiFi signals, recognize physical activities from smartphone sensors and 3D skeleton data, detect hand gestures from accelerometer signals, and so on. Program your own ensemble learning methods and use Multi-View Stacking to fuse signals from heterogeneous data sources. Use unsupervised learning algorithms to discover criminal behavioral patterns. Build deep learning neural networks with TensorFlow and Keras to classify muscle activity from electromyography signals and Convolutional Neural Networks to detect smiles in images. Evaluate the performance of

your models in traditional and multi-user settings. Build anomaly detection models such as Isolation Forests and autoencoders to detect abnormal fish behaviors. This book is intended for undergraduate/graduate students and researchers from ubiquitous computing, behavioral ecology, psychology, e-health, and other disciplines who want to learn the basics of machine learning and deep learning and for the more experienced individuals who want to apply machine learning to analyze behavioral data.

Contributions Presented at the 16th UK Workshop on Computational Intelligence, September 7–9, 2016, Lancaster, UK Springer

The 7th Paci?c Asia Conference on Knowledge Discovery and Data Mining (PAKDD) was held from April 30 to May 2, 2003 in the Convention and Ex- bition Center (COEX), Seoul, Korea. The PAKDD conference is a major forum for academic researchers and industry practitioners in the Paci?c Asia region to share original research results and development experiences from di?erent KDD-related areas such as data mining, data warehousing, machine learning, databases, statistics, knowledge acquisition and discovery, data visualization, and knowledge-based systems. The conference was organized by the Advanced Information Technology Research Center (AITrc) at KAIST and the Statistical Research Center for Complex Systems (SRCCS) at Seoul National University. It was sponsored by the Korean Datamining Society (KDMS), the Korea Inf- mation Science Society (KISS), the United States Air Force O?ce of Scienti?c Research, the Asian O?ce of Aerospace Research & Development, and KAIST. It was held with cooperation from ACM's Special Group on Knowledge Dis- very and Data Mining (SIGKDD).

Machine Learning and Knowledge Discovery in Databases Springer Nature

Much of the data available today is unstructured and text-heavy, making it challenging for analysts to apply their usual data wrangling and visualization tools. With this practical book, you'll explore text-mining techniques with tidytext, a package that authors Julia Silge and David Robinson developed using the tidy principles behind R packages like ggraph and dplyr. You'll learn how tidytext and other tidy tools in R can make text analysis easier and more effective. The authors demonstrate how treating text as data frames enables you to manipulate, summarize, and visualize characteristics of text. You'll also learn how to integrate natural language processing (NLP) into effective workflows. Practical code examples and data explorations will help you generate real insights from literature, news, and social media. Learn how to apply the tidy text format to NLP Use sentiment analysis to mine the emotional content of text Identify a document's most important terms with frequency measurements Explore relationships and connections between words with the ggraph and widyr packages Convert back and forth between R's tidy and non-tidy text formats Use topic modeling to classify document collections into natural groups Examine case studies that compare Twitter archives, dig into NASA metadata, and analyze thousands of Usenet messages

Computational Statistics with R Elsevier

Now in its second edition, this book focuses on practical algorithms for mining data from even the largest datasets.

Practical Aspects of Declarative Languages Lulu.com

Managing and Mining Graph Data is a comprehensive survey book in graph management and mining. It contains extensive surveys on a variety of important graph topics such as graph languages, indexing, clustering, data generation, pattern mining, classification, keyword search, pattern matching, and privacy. It also studies a number of domain-specific scenarios such as stream mining, web graphs, social networks, chemical and biological data. The chapters are written by well known researchers in the field, and provide a broad perspective of the area. This is the first comprehensive survey book in the emerging topic of graph data processing. Managing and Mining Graph Data is designed for a varied audience composed of professors, researchers and practitioners in industry. This volume is also suitable as a reference book for advanced-level database students in computer science and engineering.

Mining Complex Networks CRC Press

What does the Web look like? How can we find patterns, communities, outliers, in a social network? Which are the most central nodes in a network? These are the questions that motivate this work. Networks and graphs appear in many diverse settings, for example in social networks, computer-communication networks (intrusion detection, traffic management), protein-protein interaction networks in biology, document-text bipartite graphs in text retrieval, person-account graphs in financial fraud detection, and others. In this work, first we list several surprising patterns that real graphs tend to follow. Then we give a detailed list of generators that try to mirror these patterns. Generators are important, because they can help with "what if" scenarios, extrapolations, and anonymization. Then we provide a list of powerful tools for graph analysis, and specifically spectral methods (Singular Value Decomposition (SVD)), tensors, and case studies like the famous "pageRank" algorithm and the "HITS" algorithm for ranking web search results. Finally, we conclude with a survey of tools and observations from related fields like sociology, which provide complementary viewpoints. Table of Contents: Introduction / Patterns in Static Graphs / Patterns in Evolving Graphs / Patterns in Weighted Graphs / Discussion: The Structure of Specific Graphs / Discussion: Power Laws and Deviations / Summary of Patterns / Graph Generators / Preferential Attachment and Variants / Incorporating Geographical Information / The RMat / Graph Generation by Kronecker Multiplication / Summary and Practitioner's Guide / SVD, Random Walks, and Tensors / Tensors / Community Detection / Influence/Virus Propagation and Immunization / Case Studies / Social Networks / Other Related Work / Conclusions

Handbook of Statistics CRC Press

This book constitutes the thoroughly refereed post-proceedings of the 16th International Conference on Inductive Logic Programming, ILP 2006, held in Santiago de Compostela, Spain, in August 2006. The papers address all current topics in inductive logic programming, ranging from theoretical and methodological issues to advanced applications.

Managing and Mining Graph Data Cambridge University Press

Practical Graph Mining with RCRC Press