Clustering

If you ally craving such a referred **clustering** ebook that will manage to pay for you worth, get the categorically best seller from us currently from several preferred authors. If you desire to funny books, lots of novels, tale, jokes, and more fictions collections are as a consequence launched, from best seller to one of the most current released.

You may not be perplexed to enjoy all book collections clustering that we will totally offer. It is not on the costs. Its virtually what you infatuation currently. This clustering, as one of the most working sellers here will definitely be accompanied by the best options to review.

Data Clustering - Charu C. Aggarwal 2018-09-03
Research on the problem of clustering tends to be fragmented across the pattern recognition, database, data mining, and machine learning communities. Addressing this problem in a unified way, Data Clustering: Algorithms and Applications provides complete coverage of the entire area of clustering, from basic methods to more refined and complex data clustering approaches. It pays special attention to recent issues in graphs, social networks, and other domains. The book focuses on three primary aspects of data clustering: Methods, describing key techniques commonly used for clustering, such as feature selection, agglomerative clustering, partitional clustering, density-based clustering, probabilistic clustering, grid-based clustering, spectral clustering, and nonnegative matrix factorization Domains, covering methods used for different domains of data, such as categorical data, text data, multimedia data, graph data, biological data, stream data, uncertain data, time series clustering, high-dimensional clustering, and big data Variations and Insights, discussing important variations of the clustering process, such as semisupervised clustering, interactive clustering, multiview clustering, cluster ensembles, and cluster validation In this book, top researchers from around the world explore the characteristics of clustering problems in a variety of application areas. They also explain how to glean detailed insight from the clustering process—including how to verify the quality of the underlying clusters—through supervision, human intervention, or the automated generation of alternative clusters.

Reverse Clustering - Jan W. Owsiński 2021-03-03
This book presents a new perspective on and a new approach to a wide spectrum of situations, related to data analysis, actually, a kind of a new paradigm. Namely, for a given data set and its partition, whose origins may be of any kind, the authors try to reconstruct this partition on the basis of the data set given, using very broadly conceived clustering procedure. The main advantages of this new paradigm concern the substantive aspects of the particular cases considered, mainly in view of the variety of interpretations, which can be assumed in the framework of the paradigm. Due to the novel problem formulation and the flexibility in the interpretations of this problem and its components, the domains, which are encompassed (or at least affected) by the potential use of the paradigm, include cluster analysis, classification, outlier detection, feature selection, and even factor analysis as well as geometry of the data set. The book is useful for all those who look for new, nonconventional approaches to their data analysis problems.

Adaptive Resonance Theory in Social Media Data Clustering - Lei Meng 2019-04-30
Social media data contains our communication and online sharing, mirroring our daily life. This book looks at how we can use and what we can discover from such big data: Basic knowledge (data & challenges) on social media analytics Clustering as a fundamental technique for unsupervised knowledge discovery and data mining A class of neural inspired algorithms, based on adaptive resonance theory (ART), tackling challenges in big social media data

clustering Step-by-step practices of developing unsupervised machine learning algorithms for real-world applications in social media domain Adaptive Resonance Theory in Social Media Data Clustering stands on the fundamental breakthrough in cognitive and neural theory, i.e. adaptive resonance theory, which simulates how a brain processes information to perform memory, learning, recognition, and prediction. It presents initiatives on the mathematical demonstration of ART's learning mechanisms in clustering, and illustrates how to extend the base ART model to handle the complexity and characteristics of social media data and perform associative analytical tasks. Both cutting-edge research and real-world practices on machine learning and social media analytics are included in the book and if you wish to learn the answers to the following questions, this book is for you: How to process big streams of multimedia data? How to analyze social networks with heterogeneous data? How to understand a user's interests by learning from online posts and behaviors? How to create a personalized search engine by automatically indexing and searching multimodal information resources? .

Cluster Analysis for Data Mining and System Identification - János Abonyi 2007-06-22
The aim of this book is to illustrate that advanced fuzzy clustering algorithms can be used not only for partitioning of the data. It can also be used for visualization, regression, classification and time-series analysis, hence fuzzy cluster analysis is a good approach to solve complex data mining and system identification problems. This book is oriented to undergraduate and postgraduate and is well suited for teaching purposes.

Practical Guide to Cluster Analysis in R - Alboukadel Kassambara 2017-08-23
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.

Cluster Analysis - Brian S. Everitt 2001
Cluster analysis is a general term for a wide range of numerical methods used to examine multivariate data with a view to uncovering or discovering groups or clusters of homogeneous observations. This volume introduces the possibilities and limitations of clustering for research workers, as well as statisticians and graduate students in a variety of disciplines. Covers classification and clustering, visualizing clusters, measurement of proximity, hierarchical clustering, optimization techniques, finite mixture densities as models, miscellaneous methods, and comments and guidelines. Distributed by Oxford U. Press. c. Book News Inc.

Data Clustering - 2022-08-17
In view of the considerable applications of data clustering techniques in various fields, such as engineering, artificial intelligence, machine learning, clinical medicine, biology, ecology, disease diagnosis, and business marketing, many data clustering algorithms and methods have been developed to deal with complicated data. These techniques include supervised learning methods and unsupervised learning methods

such as density-based clustering, K-means clustering, and K-nearest neighbor clustering. This book reviews recently developed data clustering techniques and algorithms and discusses the development of data clustering, including measures of similarity or dissimilarity for data clustering, data clustering algorithms, assessment of clustering algorithms, and data clustering methods recently developed for insurance, psychology, pattern recognition, and survey data.

Clustering Methodology for Symbolic Data - Lynne Billard 2019-08-12
Covers everything readers need to know about clustering methodology for symbolic data—including new methods and headings—while providing a focus on multi-valued list data, interval data and histogram data This book presents all of the latest developments in the field of clustering methodology for symbolic data—paying special attention to the classification methodology for multi-valued list, interval-valued and histogram-valued data methodology, along with numerous worked examples. The book also offers an expansive discussion of data management techniques showing how to manage the large complex dataset into more manageable datasets ready for analyses. Filled with examples, tables, figures, and case studies, Clustering Methodology for Symbolic Data begins by offering chapters on data management, distance measures, general clustering techniques, partitioning, divisive clustering, and agglomerative and pyramid clustering. Provides new classification methodologies for histogram valued data reaching across many fields in data science Demonstrates how to manage a large complex dataset into manageable datasets ready for analysis Features very large contemporary datasets such as multi-valued list data, interval-valued data, and histogram-valued data Considers classification models by dynamical clustering Features a supporting website hosting relevant data sets Clustering Methodology for Symbolic Data will appeal to practitioners of symbolic data analysis, such as statisticians and economists within the public sectors. It will also be of interest to postgraduate students of, and researchers within, web mining, text mining and bioengineering.

Introduction to Clustering Large and High-Dimensional Data - Jacob Kogan 2007
Focuses on a few of the important clustering algorithms in the context of information retrieval.

Machine Learning with Clustering - Artem Kovera 2017-10-24
There are four major tasks for clustering: Making simplification for further data processing. In this case, the data is split into different groups which then are processed individually. In business, for instance, we can find different groups of customers sharing some similar features using cluster analysis. Then, we can use this information to develop different marketing strategies and apply them to all these separate groups of customers. Or, we can cluster a marketplace in a specific niche to find what kinds of products are selling better than other ones to make a decision what kind of products to produce. Usually, clustering is one of the first techniques that help explore a dataset we are going to work with to get some sense of the structure of the data. Compression of the data. We can implement cluster analysis on a giant data set. Then from each cluster, we can pick just several items. In this case, we usually lose much less information than in the case where we pick data points without preceding clustering. Clustering algorithms are being used to compress not only large data sets but also relatively small objects like images. Picking out unusual data points from the dataset. This procedure is done, for example, for the detection of fraudulent transactions with credit cards. In medicine, similar procedures can be used, for example, to identify new forms of illnesses. Building the hierarchy of objects. This is implemented for classification of biological organisms. It is also applied, for example, in search engines to group different text documents inside the search engines' datasets. In an introductory chapter, you will find: Different types of machine learning; Features in datasets; Dimensionality of datasets; The 'curse' of dimensionality; Dealing with underfitting and overfitting In the following chapters, we will implement these concepts in practice, working with clustering algorithms. This book provides detailed explanations of several widely-used clustering approaches with visual

representations: Hierarchical agglomerative clustering; K-means; DBSCAN; Neural network-based clustering You will learn different strengths and weaknesses of these algorithms as well as the practical strategies to overcome the weaknesses. In addition, we will briefly touch upon some other clustering methods. The examples of the algorithms are presented in Python 3. We will work with several datasets, including the ones based on real-world data. We will be primarily working with the Scikit-learn and SciPy libraries. But our neural network for clustering, we will build basically from scratch, just by using NumPy arrays.

Unsupervised Machine Learning for Clustering in Political and Social Research - Philip D. Waggoner 2021-01-28
In the age of data-driven problem-solving, applying sophisticated computational tools for explaining substantive phenomena is a valuable skill. Yet, application of methods assumes an understanding of the data, structure, and patterns that influence the broader research program. This Element offers researchers and teachers an introduction to clustering, which is a prominent class of unsupervised machine learning for exploring and understanding latent, non-random structure in data. A suite of widely used clustering techniques is covered in this Element, in addition to R code and real data to facilitate interaction with the concepts. Upon setting the stage for clustering, the following algorithms are detailed: agglomerative hierarchical clustering, k-means clustering, Gaussian mixture models, and at a higher-level, fuzzy C-means clustering, DBSCAN, and partitioning around medoids (k-medoids) clustering.

Clustering Techniques for Image Segmentation - Fasahat Ullah Siddiqui 2021-10-29
This book presents the workings of major clustering techniques along with their advantages and shortcomings. After introducing the topic, the authors illustrate their modified version that avoids those shortcomings. The book then introduces four modified clustering techniques, namely the Optimized K-Means (OKM), Enhanced Moving K-Means-1(EMKM-1), Enhanced Moving K-Means-2(EMKM-2), and Outlier Rejection Fuzzy C-Means (ORFCM). The authors show how the OKM technique can differentiate the empty and zero variance cluster, and the data assignment procedure of the K-mean clustering technique is redesigned. They then show how the EMKM-1 and EMKM-2 techniques reform the data-transferring concept of the Adaptive Moving K-Means (AMKM) to avoid the centroid trapping problem. And that the ORFCM technique uses the adaptable membership function to moderate the outlier effects on the Fuzzy C-meaning clustering technique. This book also covers the working steps and codings of quantitative analysis methods. The results highlight that the modified clustering techniques generate more homogenous regions in an image with better shape and sharp edge preservation. Showcases major clustering techniques, detailing their advantages and shortcomings; Includes several methods for evaluating the performance of segmentation techniques; Presents several applications including medical diagnosis systems, satellite imaging systems, and biometric systems.

Similarity-Based Clustering - Thomas Villmann 2009-06-02
This book is the outcome of the Dagstuhl Seminar on "Similarity-Based Clustering" held at Dagstuhl Castle, Germany, in Spring 2007. In three chapters, the three fundamental aspects of a theoretical background, the representation of data and their connection to algorithms, and particular challenging applications are considered. Topics discussed concern a theoretical investigation and foundation of prototype based learning algorithms, the development and extension of models to directions such as general data structures and the application for the domain of medicine and biology. Similarity based methods find widespread applications in diverse application domains, including biomedical problems, but also in remote sensing, geoscience or other technical domains. The presentations give a good overview about important research results in similarity-based learning, whereby the character of overview articles with references to correlated research articles makes the contributions particularly suited for a first reading concerning these topics.

Model-Based Clustering and Classification for Data Science - Charles Bouveyron

2019-09-30
Cluster analysis finds groups in data automatically. Most methods have been heuristic and leave open such central questions as: how many clusters are there? Which method should I use? How should I handle outliers? Classification assigns new observations to groups given previously classified observations, and also has open questions about parameter tuning, robustness and uncertainty assessment. This book frames cluster analysis and classification in terms of statistical models, thus yielding principled estimation, testing and prediction methods, and sound answers to the central questions. It builds the basic ideas in an accessible but rigorous way, with extensive data examples and R code; describes modern approaches to high-dimensional data and networks; and explains such recent advances as Bayesian regularization, non-Gaussian model-based clustering, cluster merging, variable selection, semi-supervised and robust classification, clustering of functional data, text and images, and co-clustering. Written for advanced undergraduates in data science, as well as researchers and practitioners, it assumes basic knowledge of multivariate calculus, linear algebra, probability and statistics.

Clustering and Fuzzy Techniques: Theory, Implementation and Applications - Hizir 2003

An Introduction to Clustering with R - Paolo Giordani 2020-08-27
The purpose of this book is to thoroughly prepare the reader for applied research in clustering. Cluster analysis comprises a class of statistical techniques for classifying multivariate data into groups or clusters based on their similar features. Clustering is nowadays widely used in several domains of research, such as social sciences, psychology, and marketing, highlighting its multidisciplinary nature. This book provides an accessible and comprehensive introduction to clustering and offers practical guidelines for applying clustering tools by carefully chosen real-life datasets and extensive data analyses. The procedures addressed in this book include traditional hard clustering methods and up-to-date developments in soft clustering. Attention is paid to practical examples and

applications through the open source statistical software R. Commented R code and output for conducting, step by step, complete cluster analyses are available. The book is intended for researchers interested in applying clustering methods. Basic notions on theoretical issues and on R are provided so that professionals as well as novices with little or no background in the subject will benefit from the book.
Algorithms for Clustering Data - Anil K. Jain 1988

Clustering - Boris Mirkin 2012-10-17
Often considered more of an art than a science, books on clustering have been dominated by learning through example with techniques chosen almost through trial and error. Even the two most popular, and most related, clustering methods—K-Means for partitioning and Ward's method for hierarchical clustering—have lacked the theoretical underpinning required to establish a firm relationship between the two methods and relevant interpretation aids. Other approaches, such as spectral clustering or consensus clustering, are considered absolutely unrelated to each other or to the two above mentioned methods. Clustering: A Data Recovery Approach, Second Edition presents a unified modeling approach for the most popular clustering methods: the K-Means and hierarchical techniques, especially for divisive clustering. It significantly expands coverage of the mathematics of data recovery, and includes a new chapter covering more recent popular network clustering approaches—spectral, modularity and uniform, additive, and consensus—treated within the same data recovery approach. Another added chapter covers cluster validation and interpretation, including recent developments for ontology-driven interpretation of clusters. Altogether, the insertions added a hundred pages to the book, even in spite of the fact that fragments unrelated to the main topics were removed. Illustrated using a set of small real-world datasets and more than a hundred examples, the book is oriented towards students, practitioners, and theoreticians of cluster analysis. Covering topics that are beyond the scope of most texts, the author's explanations of data recovery methods, theory-based advice, pre- and post-processing

issues and his clear, practical instructions for real-world data mining make this book ideally suited for teaching, self-study, and professional reference.

An Occupational Clustering System and Curriculum - John E. Taylor 1972

Fuzzy Cluster Analysis - Frank Höppner 1999-07-09
Provides a timely and important introduction to fuzzy cluster analysis, its methods and areas of application, systematically describing different fuzzy clustering techniques so the user may choose methods appropriate for his problem. It provides a very thorough overview of the subject and covers classification, image recognition, data analysis and rule generation. The application examples are highly relevant and illustrative, and the use of the techniques are justified and well thought-out. Features include: * Sections on inducing fuzzy if-then rules by fuzzy clustering and non-alternating optimization fuzzy clustering algorithms * Discussion of solid fuzzy clustering techniques like the fuzzy c-means, the Gustafson-Kessel and the Gath-and-Geva algorithm for classification problems * Focus on linear and shell clustering techniques used for detecting contours in image analysis * Accompanying software and data sets pertaining to the examples presented, enabling the reader to learn through experimentation * Examination of the difficulties involved in evaluating the results of fuzzy cluster analysis and of determining the number of clusters with analysis of global and local validity measures This is one of the most comprehensive books on fuzzy clustering and will be welcomed by computer scientists, engineers and mathematicians in industry and research who are concerned with different methods, data analysis, pattern recognition or image processing. It will also give graduate students in computer science, mathematics or statistics a valuable overview.
UNSUPERVISED CLUSTERING CATEGORICAL DATA USING EVOLUTIONARY OPTIMIZATION TECHNIQUES - Dr. G. Surya Narayana

Large-Scale Group Decision-Making - Su-Min Yu 2022
This book explores clustering operations in the context of social networks and consensus-reaching paths that take into account non-cooperative behaviors. This book focuses on the two key issues in large-scale group decision-making: clustering and consensus building. Clustering aims to reduce the dimension of a large group. Consensus reaching requires that the divergent individual opinions of the decision makers converge to the group opinion. This book emphasizes the similarity of opinions and social relationships as important measurement attributes of clustering, which makes it different from traditional clustering methods with single attribute to divide the original large group without requiring a combination of the above two attributes. The proposed consensus models focus on the treatment of non-cooperative behaviors in the consensus-reaching process and explores the influence of trust loss on the consensus-reaching process.The logic behind is as follows: firstly, a clustering algorithm is adopted to reduce the dimension of decision-makers, and then, based on the clusters opinions obtained, a consensus-reaching process is carried out to obtain a decision result acceptable to the majority of decision-makers. Graduates and researchers in the fields of management science, computer science, information management, engineering technology, etc., who are interested in large-scale group decision-making and consensus building are potential audience of this book. It helps readers to have a deeper and more comprehensive understanding of clustering analysis and consensus building in large-scale group decision-making. .
Clustering - Rui Xu 2008-11-03
This is the first book to take a truly comprehensive look at clustering. It begins with an introduction to cluster analysis and goes on to explore: proximity measures; hierarchical clustering; partition clustering; neural network-based clustering; kernel-based clustering; sequential data clustering; large-scale data clustering; data visualization and high-dimensional data clustering; and cluster validation. The authors assume no previous background in clustering and their generous inclusion of examples and references help make the subject matter comprehensible for readers of varying levels and backgrounds.
Clustering Challenges in Biological

Networks -

Introduction to Microsoft Windows NT Cluster Server - Raj Rajagopal 1999-12-15 Mastering cluster technology-the linking of servers-is becoming increasingly important for application and system programmers and network designers, administrators, and managers. With Microsoft's Windows NT cluster server being the first to tie cluster technology with a major operating system, it appears destined to take a leadership position in the industry. Introduction to Microsoft Windows NT Cluster Server provides all you need to know to develop your abilities for this essential technology. The author provides both introductory and advanced material focused on the three basic functions: fault tolerant computing (failover), load balancing, and centralized administration and monitoring. He guides the reader from the basics of cluster servers, through Microsoft's cluster server set-up, communication, programming, and administration. Written for professionals who are familiar with the Windows NT operating system and have programming experience, Introduction to Microsoft Windows NT Cluster Server contains information instrumental in helping you achieve zero downtime.
Recent Advances in Hybrid Metaheuristics for Data Clustering - Sourav De 2020-08-24 An authoritative guide to an in-depth analysis of various state-of-the-art data clustering approaches using a range of computational intelligence techniques Recent Advances in Hybrid Metaheuristics for Data Clustering offers a guide to the fundamentals of various metaheuristics and their application to data clustering. Metaheuristics are designed to tackle complex clustering problems where classical clustering algorithms have failed to be either effective or efficient. The authors—noted experts on the topic—provide a text that can aid in the design and development of hybrid metaheuristics to be applied to data clustering. The book includes performance analysis of the hybrid metaheuristics in relationship to their conventional counterparts. In addition to providing a review of data clustering, the authors include in-depth analysis of different optimization algorithms. The text offers a step-by-step guide in the build-up of hybrid metaheuristics and to enhance comprehension. In addition, the book contains a range of real-life case studies and their applications. This important text: Includes performance analysis of the hybrid metaheuristics as related to their conventional counterparts Offers an in-depth analysis of a range of optimization algorithms Highlights a review of data clustering Contains a detailed overview of different standard metaheuristics in current use Presents a step-by-step guide to the build-up of hybrid metaheuristics Offers real-life case studies and applications Written for researchers, students and academics in computer science, mathematics, and engineering, Recent Advances in Hybrid Metaheuristics for Data Clustering provides a text that explores the current data clustering approaches using a range of computational intelligence techniques.
Advances in Fuzzy Clustering and its Applications - Jose Valente de Oliveira 2007-06-13 A comprehensive, coherent, and in depth presentation of the state of the art in fuzzy clustering. Fuzzy clustering is now a mature and vibrant area of research with highly innovative advanced applications. Encapsulating this through presenting a careful selection of research contributions, this book addresses timely and relevant concepts and methods, whilst identifying major challenges and recent developments in the area. Split into five clear sections, Fundamentals, Visualization, Algorithms and Computational Aspects, Real-Time and Dynamic Clustering, and Applications and Case Studies, the book covers a wealth of novel, original and fully updated material, and in particular offers: a focus on the algorithmic and computational augmentations of fuzzy clustering and its effectiveness in handling high dimensional problems, distributed problem solving and uncertainty management. presentations of the important and relevant phases of cluster design, including the role of information granules, fuzzy sets in the realization of human-centricity facet of data analysis, as well as system modelling demonstrations of how the results facilitate further detailed development of models, and enhance interpretation aspects a carefully

organized illustrative series of applications and case studies in which fuzzy clustering plays a pivotal role This book will be of key interest to engineers associated with fuzzy control, bioinformatics, data mining, image processing, and pattern recognition, while computer engineers, students and researchers, in most engineering disciplines, will find this an invaluable resource and research tool.

Advances in K-means Clustering - Junjie Wu 2012-07-09
Nearly everyone knows K-means algorithm in the fields of data mining and business intelligence. But the ever-emerging data with extremely complicated characteristics bring new challenges to this "old" algorithm. This book addresses these challenges and makes novel contributions in establishing theoretical frameworks for K-means distances and K-means based consensus clustering, identifying the "dangerous" uniform effect and zero-value dilemma of K-means, adapting right measures for cluster validity, and integrating K-means with SVMs for rare class analysis. This book not only enriches the clustering and optimization theories, but also provides good guidance for the practical use of K-means, especially for important tasks such as network intrusion detection and credit fraud prediction. The thesis on which this book is based has won the "2010 National Excellent Doctoral Dissertation Award", the highest honor for not more than 100 PhD theses per year in China.

Proceedings of the 7th International Conference on Clustering Aspects of Nuclear Structure and Dynamics - M. Korolija 2000
In the past three decades our understanding of the clustering behavior of nucleons in both nuclear structure and nuclear dynamics has evolved considerably. Moreover, the notion of the cluster has made its way into a number of scientific disciplines. This book provides an overview of the current understanding of clustering phenomena in nuclear structure and nuclear dynamics. The topics covered include: fundamental aspects of nuclear clustering, models of nucleon clusterization, clustering aspects of nuclear structure, selected topics on clustering aspects in medium- and high-energy nucleus-nucleus collisions.

Data Clustering - Guojun Gan 2007-07-12
Reference and compendium of algorithms for pattern recognition, data mining and statistical computing.

Evolutionary Data Clustering: Algorithms and Applications - Ibrahim Aljarah 2021-02-20
This book provides an in-depth analysis of the current evolutionary clustering techniques. It discusses the most highly regarded methods for data clustering. The book provides literature reviews about single objective and multi-objective evolutionary clustering algorithms. In addition, the book provides a comprehensive review of the fitness functions and evaluation measures that are used in most of evolutionary clustering algorithms. Furthermore, it provides a conceptual analysis including definition, validation and quality measures, applications, and implementations for data clustering using classical and modern nature-inspired techniques. It features a range of proven and recent nature-inspired algorithms used to data clustering, including particle swarm optimization, ant colony optimization, grey wolf optimizer, salp swarm algorithm, multi-verse optimizer, Harris hawks optimization, beta-hill climbing optimization. The book also covers applications of evolutionary data clustering in diverse fields such as image segmentation, medical applications, and pavement infrastructure asset management.

Data Clustering: Theory, Algorithms, and Applications, Second Edition - Guojun Gan 2020-11-10
Data clustering, also known as cluster analysis, is an unsupervised process that divides a set of objects into homogeneous groups. Since the publication of the first edition of this monograph in 2007, development in the area has exploded, especially in clustering algorithms for big data and open-source software for cluster analysis. This second edition reflects these new developments, covers the basics of data clustering, includes a list of popular clustering algorithms, and provides program code that helps users implement clustering algorithms. Data Clustering: Theory, Algorithms and Applications, Second Edition will be of interest to researchers, practitioners, and data scientists as well as undergraduate and graduate students.
Algorithms for Fuzzy Clustering - Sadaaki

Miyamoto 2008-04-15

Recently many researchers are working on cluster analysis as a main tool for exploratory data analysis and data mining. A notable feature is that specialists in di?erent ?elds of sciences are considering the tool of data clustering to be useful. A major reason is that clustering algorithms and software are ?exible in thesensethatdi?erentmathematicalframeworksar eemployedinthealgorithms and a user can select a suitable method according to his application. Moreover clusteringalgorithmshavedi?erentoutputsranging fromtheolddendrogramsof agglomerativeclustering to more recent self-organizingmaps. Thus, a researcher or user can choose an appropriate output suited to his purpose,which is another ?exibility of the methods of clustering. An old and still most popular method is the K-means which use K cluster centers. A group of data is gathered around a cluster center and thus forms a cluster. The main subject of this book is the fuzzy c-means proposed by Dunn and Bezdek and their variations including recent studies. A main reasonwhy we concentrate on fuzzy c-means is that most methodology and application studies infuzzy clusteringusefuzzy c-means,andfuzzy c-meansshouldbe consideredto beamajortechniqueofclusteringingeneral,regardl esswhetheroneisinterested in fuzzy methods or not. Moreover recent advances in clustering techniques are rapid and we requirea new textbook that includes recent algorithms.We should also note that several books have recently been published but the contents do not include some methods studied herein.

Cluster - Piers Anthony 2014-04-01

First in the sci-fi series "packed with exotic beings, ancient secrets, and futuristic worlds" from the New York Times–bestselling author (The Portalist). As Cluster opens, the alien envoy Pnotl of Sphere Knyfh seeks help from Sphere Sol in a shared galactic-level crisis: Galaxy Andromeda has discovered the secret of energy transfer and intends to use it to steal the basic energy of the Milky Way Galaxy. Knyfh offers the secret of aura transfer on the understanding that Sphere Sol will spread the technology to help create a galactic coalition to find and defeat agents of Andromeda. Sol's highest-Kirlian individual is Flint, a green-skinned native of Outworld, who has a Kirlian aura of two hundred, an eidetic memory (useful for memorizing the complex equations of Kirlian transfer that he will need to communicate to other spheres). He has extraordinary intelligence and is highly adaptable. His mission is complicated, however, by the fact that he is pursued everywhere by a very high Kirlian female Andromedan agent and, somehow, the Andromedans are able to detect and trace Kirlian transfers. Flint embarks upon several missions to bring transfer technology to neighboring spheres, inhabiting various alien forms. His efforts are successful despite attacks and sabotage by the Andromedan agent. Through the conflict, however, the mutual attraction of their two vastly superior auras begins to undermine their individual loyalty to their own Spheres. Flint and a group of other entities recover the information that will allow them to detect and trace transfers, and a member of the group is revealed as the Andromedan agent. One result is the catastrophic destruction of the local habitat. Flint and his nemesis are transferred into alien Mintakan bodies to survive. Choosing to leave things with parity between their two galaxies, Flint and the Andromedan mate and remain together until their auras fade (which happens rapidly, since their physical bodies have been destroyed).

Data Clustering - Charu C. Aggarwal 2013-08-21

Research on the problem of clustering tends to be fragmented across the pattern recognition, database, data mining, and machine learning communities. Addressing this problem in a unified way, Data Clustering: Algorithms and Applications provides complete coverage of the entire area of clustering, from basic methods to more refined and complex data clustering approaches. It pays special attention to recent issues in graphs, social networks, and other domains. The book focuses on three primary aspects of data clustering: Methods, describing key techniques commonly used for clustering, such as feature selection, agglomerative clustering, partitional clustering, density-based clustering, probabilistic clustering, grid-based clustering, spectral clustering, and nonnegative

matrix factorization Domains, covering methods used for different domains of data, such as categorical data, text data, multimedia data, graph data, biological data, stream data, uncertain data, time series clustering, high-dimensional clustering, and big data Variations and Insights, discussing important variations of the clustering process, such as semisupervised clustering, interactive clustering, multiview clustering, cluster ensembles, and cluster validation In this book, top researchers from around the world explore the characteristics of clustering problems in a variety of application areas. They also explain how to glean detailed insight from the clustering process—including how to verify the quality of the underlying clusters—through supervision, human intervention, or the automated generation of alternative clusters.

Co-Clustering - Gérard Govaert 2013-12-11 Cluster or co-cluster analyses are important tools in a variety ofscientific areas. The introduction of this book presents a state ofthe art of already well-established, as well as more recent methodsof co-clustering. The authors mainly deal with the two-modepartitioning under different approaches, but pay particularattention to a probabilistic approach. Chapter 1 concerns clustering in general and the model-basedclustering in particular. The authors briefly review the classicalclustering methods and focus on the mixture model. They present anddiscuss the use of different mixtures adapted to different types ofdata. The algorithms used are described and related works withdifferent classical methods are presented and commented upon. Thischapter is useful in tackling the problem of co-clustering under the mixture approach. Chapter 2 is devoted tothe latent block model proposed in the mixture approach context.The authors discuss this model in detail and present its interestregarding co-clustering. Various algorithms are presented in ageneral context. Chapter 3 focuses on binary and categorical data.It presents, in detail, the appropriated latent block mixturemodels. Variants of these models and algorithms are presented andillustrated using examples. Chapter 4 focuses on contingency data.Mutual information, phi-squared and model-based co-clustering arestudied. Models, algorithms and

connections among differentapproaches are described and illustrated. Chapter 5 presents thecase of continuous data. In the same way, the different approachesused in the previous chapters are extended to this situation. Contents 1. Cluster Analysis. 2. Model-Based Co-Clustering. 3. Co-Clustering of Binary and Categorical Data. 4. Co-Clustering of Contingency Tables. 5. Co-Clustering of Continuous Data. About the Authors Gérard Govaert is Professor at the University of Technologyof Compiègne, France. He is also a member of the CNRSLaboratory Heudiasyc (Heuristic and diagnostic of complex systems).His research interests include latent structure modeling, modelselection, model-based cluster analysis, block clustering andstatistical pattern recognition. He is one of the authors of theMIXMOD (MIXtureMODelling) software. Mohamed Nadif is Professor at the University of Paris-Descartes,France, where he is a member of LIPADE (Paris Descartes computerscience laboratory) in the Mathematics and Computer Sciencedepartment. His research interests include machine learning, datamining, model-based cluster analysis, co-clustering, factorizationand data analysis. Cluster Analysis is an important tool in a variety of scientificareas. Chapter 1 briefly presents a state of the art of alreadywell-established as well more recent methods. The hierarchical,partitioning and fuzzy approaches will be discussed amongst others.The authors review the difficulty of these classical methods intackling the high dimensionality, sparsity and scalability. Chapter2 discusses the interests of coclustering, presenting differentapproaches and defining a co-cluster. The authors focus onco-clustering as a simultaneous clustering and discuss the cases ofbinary, continuous and co-occurrence data. The criteria andalgorithms are described and illustrated on simulated and realdata. Chapter 3 considers co-clustering as a model-basedco-clustering. A latent block model is defined for different kindsof data. The estimation of parameters and co-clustering is tackledunder two approaches: maximum likelihood and classification maximumlikelihood. Hard and soft algorithms are described and applied onsimulated and real data. Chapter 4 considers co-clustering as amatrix

approximation. The trifactorization approach is consideredand algorithms based on update rules are described. Links withnumerical and probabilistic approaches are established. Acombination of algorithms are proposed and evaluated on simulatedand real data. Chapter 5 considers a co-clustering or bi-clusteringas the search for coherent co-clusters in biological terms or theextraction of co-clusters under conditions. Classical algorithmswill be described and evaluated on simulated and real data.Different indices to evaluate the quality of coclusters are notedand used in numerical experiments.

Projection-Based Clustering through Self-Organization and Swarm Intelligence - Michael Christoph Thrun 2018-01-09 This book is published open access under a CC BY 4.0 license. It covers aspects of unsupervised machine learning used for knowledge discovery in data science and introduces a data-driven approach to cluster analysis, the Databionic swarm (DBS). DBS consists of the 3D landscape visualization and clustering of data. The 3D landscape enables 3D printing of high-dimensional data structures. The clustering and number of clusters or an absence of cluster structure are verified by the 3D landscape at a glance. DBS is the first swarm-based technique that shows emergent properties while exploiting concepts of swarm intelligence, self-organization and the Nash equilibrium concept from game theory. It results in the elimination of a global objective function and the setting of parameters. By downloading the R package DBS can be applied to data drawn from diverse research fields and used even by non-professionals in the field of data mining.

Clustering and Information Retrieval - Weili Wu 2003-11-30 Clustering is an important technique for discovering relatively dense sub-regions or sub-spaces of a multi-dimension data distribution. Clus tering has been used in information retrieval for many different purposes, such as query expansion, document grouping, document indexing, and visualization of search results. In this book, we address issues of cluster ing algorithms, evaluation methodologies, applications, and architectures for information retrieval. The first two chapters discuss clustering algorithms. The chapter from Baeza-Yates et al. describes a clustering method for a general metric space which is a common model of data relevant to information retrieval. The chapter by Guha, Rastogi, and Shim presents a survey as well as detailed discussion of two clustering algorithms: CURE and ROCK for numeric data and categorical data respectively. Evaluation methodologies are addressed in the next two chapters. Ertoz et al. demonstrate the use of text retrieval benchmarks, such as TRECS, to evaluate clustering algorithms. He et al. provide objective measures of clustering quality in their chapter. Applications of clustering methods to information retrieval is ad dressed in the next four chapters. Chu et al. and Noel et al. explore feature selection using word stems, phrases, and link associations for document clustering and indexing. Wen et al. and Sung et al. discuss applications of clustering to user queries and data cleansing. Finally, we consider the problem of designing architectures for infor mation retrieval. Crichton, Hughes, and Kelly elaborate on the devel opment of a scientific data system architecture for information retrieval.

Clustering Methods for Big Data Analytics - Olfa Nasraoui 2018-10-27 This book highlights the state of the art and recent advances in Big Data clustering methods and their innovative applications in contemporary AI-driven systems. The book chapters discuss Deep Learning for Clustering, Blockchain data clustering, Cybersecurity applications such as insider threat detection, scalable distributed clustering methods for massive volumes of data; clustering Big Data Streams such as streams generated by the confluence of Internet of Things, digital and mobile health, human-robot interaction, and social networks; Spark-based Big Data clustering using Particle Swarm Optimization; and Tensor-based clustering for Web graphs, sensor streams, and social networks. The chapters in the book include a balanced coverage of big data clustering theory, methods, tools, frameworks, applications, representation, visualization, and clustering validation.

Recent Applications in Data Clustering - Harun Pirim 2018-08-01 Clustering has emerged as one of the more

fertile fields within data analytics, widely adopted by companies, research institutions, and educational entities as a tool to describe similar/different groups. The book Recent Applications in Data Clustering aims to provide an outlook of recent contributions to the vast clustering literature that offers useful insights within the context of modern applications for professionals, academics, and students. The book spans the domains of clustering in image analysis, lexical analysis of texts, replacement of missing values in data, temporal clustering in smart cities, comparison of artificial neural network variations, graph theoretical approaches, spectral clustering, multiview clustering, and model-based clustering in an R package. Applications of image, text, face recognition, speech (synthetic and simulated), and smart city datasets are presented.