

Frequent Pattern Mining Charu Aggarwal

Comprehensive Coverage of the Entire Area of Classification Research on the problem of classification tends to be fragmented across such areas as pattern recognition, database, data mining, and machine learning. Addressing the work of these different communities in a unified way, *Data Classification: Algorithms and Applications* explores the underlying

The fundamental mathematical tools needed to understand machine learning include linear algebra, analytic geometry, matrix decompositions, vector calculus, optimization, probability and statistics. These topics are traditionally taught in disparate courses, making it hard for data science or computer science students, or professionals, to efficiently learn the mathematics. This self-contained textbook bridges the gap between mathematical and machine learning texts, introducing the mathematical concepts with a minimum of prerequisites. It uses these concepts to derive four central machine learning methods: linear regression, principal component analysis, Gaussian mixture models and support vector machines. For students and others with a mathematical background, these derivations provide a starting point to machine learning texts. For those learning the mathematics for the first time, the methods help build intuition and practical experience with applying mathematical concepts. Every chapter includes worked examples and exercises to test understanding. Programming tutorials are offered on the book's web site.

This textbook covers the broader field of artificial intelligence. The chapters for this textbook span within three categories: *Deductive reasoning methods*: These methods start with pre-defined hypotheses and reason with them in order to arrive at logically sound conclusions. The underlying methods include search and logic-based methods. These methods are discussed in Chapters 1 through 5. *Inductive Learning Methods*: These methods start with examples and use statistical methods in order to arrive at hypotheses. Examples include regression modeling, support vector machines, neural networks, reinforcement learning, unsupervised learning, and probabilistic graphical models. These methods are discussed in Chapters 6 through 11. *Integrating Reasoning and Learning*: Chapters 11 and 12 discuss techniques for integrating reasoning and learning. Examples include the use of knowledge graphs and neuro-symbolic artificial intelligence. The primary audience for this textbook are professors and advanced-level students in computer science. It is also possible to use this textbook for the mathematics requirements for an undergraduate data science course. Professionals working in this related field many also find this textbook useful as a reference.

This book covers both classical and modern models in deep learning. The primary focus is on the theory and algorithms of deep learning. The theory and algorithms of neural networks are particularly important for understanding important concepts, so that one can understand the important design concepts of neural architectures in different applications. Why do neural networks work? When do they work better than off-the-shelf machine-learning models? When is depth useful? Why is training neural networks so hard? What are the pitfalls? The book is also rich in discussing different applications in order to give the practitioner a flavor of how neural architectures are designed for different types of problems. Applications associated with many different areas like recommender systems, machine translation, image captioning, image classification, reinforcement-learning based gaming, and text analytics are covered. The chapters of this book span three categories: *The basics of neural networks*: Many traditional machine learning models can be understood as special cases of neural networks. An emphasis is placed in the first two chapters on understanding the relationship between traditional machine learning and neural networks. *Support vector machines, linear/logistic regression, singular value*

decomposition, matrix factorization, and recommender systems are shown to be special cases of neural networks. These methods are studied together with recent feature engineering methods like word2vec. Fundamentals of neural networks: A detailed discussion of training and regularization is provided in Chapters 3 and 4. Chapters 5 and 6 present radial-basis function (RBF) networks and restricted Boltzmann machines. Advanced topics in neural networks: Chapters 7 and 8 discuss recurrent neural networks and convolutional neural networks. Several advanced topics like deep reinforcement learning, neural Turing machines, Kohonen self-organizing maps, and generative adversarial networks are introduced in Chapters 9 and 10. The book is written for graduate students, researchers, and practitioners. Numerous exercises are available along with a solution manual to aid in classroom teaching. Where possible, an application-centric view is highlighted in order to provide an understanding of the practical uses of each class of techniques.

Data Mining

Managing and Mining Sensor Data

The Textbook

An Ontology-based Approach

Data Classification

The Art and Science of Algorithms that Make Sense of Data

This comprehensive reference consists of 18 chapters from prominent researchers in the field. Each chapter is self-contained, and synthesizes one aspect of frequent pattern mining. An emphasis is placed on simplifying the content, so that students and practitioners can benefit from the book. Each chapter contains a survey describing key research on the topic, a case study and future directions. Key topics include: Pattern Growth Methods, Frequent Pattern Mining in Data Streams, Mining Graph Patterns, Big Data Frequent Pattern Mining, Algorithms for Data Clustering and more. Advanced-level students in computer science, researchers and practitioners from industry will find this book an invaluable reference.

Increasingly, human beings are sensors engaging directly with the mobile Internet. Individuals can now share real-time experiences at an unprecedented scale. Social Sensing: Building Reliable Systems on Unreliable Data looks at recent advances in the emerging field of social sensing, emphasizing the key problem faced by application designers: how to extract reliable information from data collected from largely unknown and possibly unreliable sources. The book explains how a myriad of societal applications can be derived from this massive amount of data collected and shared by average individuals. The title offers theoretical foundations to support emerging data-driven cyber-physical applications and touches on key issues such as privacy. The authors present solutions based on recent research and novel ideas that leverage techniques from cyber-physical systems, sensor networks, machine learning, data mining, and information fusion. Offers a unique interdisciplinary perspective bridging social networks, big data, cyber-physical systems, and reliability Presents novel theoretical foundations

for assured social sensing and modeling humans as sensors Includes case studies and application examples based on real data sets Supplemental material includes sample datasets and fact-finding software that implements the main algorithms described in the book

Interest in predictive analytics of big data has grown exponentially in the four years since the publication of Statistical and Machine-Learning Data Mining: Techniques for Better Predictive Modeling and Analysis of Big Data, Second Edition. In the third edition of this bestseller, the author has completely revised, reorganized, and repositioned the original chapters and produced 13 new chapters of creative and useful machine-learning data mining techniques. In sum, the 43 chapters of simple yet insightful quantitative techniques make this book unique in the field of data mining literature. What is new in the Third Edition: The current chapters have been completely rewritten. The core content has been extended with strategies and methods for problems drawn from the top predictive analytics conference and statistical modeling workshops. Adds thirteen new chapters including coverage of data science and its rise, market share estimation, share of wallet modeling without survey data, latent market segmentation, statistical regression modeling that deals with incomplete data, decile analysis assessment in terms of the predictive power of the data, and a user-friendly version of text mining, not requiring an advanced background in natural language processing (NLP). Includes SAS subroutines which can be easily converted to other languages. As in the previous edition, this book offers detailed background, discussion, and illustration of specific methods for solving the most commonly experienced problems in predictive modeling and analysis of big data. The author addresses each methodology and assigns its application to a specific type of problem. To better ground readers, the book provides an in-depth discussion of the basic methodologies of predictive modeling and analysis. While this type of overview has been attempted before, this approach offers a truly nitty-gritty, step-by-step method that both tyros and experts in the field can enjoy playing with. Covering all the main approaches in state-of-the-art machine learning research, this will set a new standard as an introductory textbook.

Web Data Mining

Data Mining: Concepts and Techniques

A Data Recovery Approach

Social Media Mining

Machine Learning for Data Streams

Theory, Algorithms, and Applications

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.

Text mining applications have experienced tremendous advances because of web 2.0 and social networking applications. Recent advances in hardware and software technology have lead to a number of unique scenarios where text mining algorithms are learned. Mining Text Data introduces an important niche in the text analytics field, and is an edited volume contributed by leading international researchers and practitioners focused on social networks & data mining. This book contains a wide swath in topics across social networks & data mining. Each chapter contains a comprehensive survey including the key research content on the topic, and the future directions of research in the field. There is a special focus on Text Embedded with Heterogeneous and Multimedia Data which makes the mining process much more challenging. A number of methods have been designed such as transfer learning and cross-lingual mining for such cases. Mining Text Data simplifies the content, so that advanced-level students, practitioners and researchers in computer science can benefit from this book. Academic and corporate libraries, as well as ACM, IEEE, and Management Science focused on information security, electronic commerce, databases, data mining, machine learning, and statistics are the primary buyers for this reference book.

Cluster analysis is an unsupervised process that divides a set of objects into homogeneous groups. This book starts with basic information on cluster analysis, including the classification of data and the corresponding similarity measures, followed by the presentation of over 50 clustering algorithms in groups according to some specific baseline methodologies such as hierarchical, center-based, and search-based methods. As a result, readers and users can easily identify an appropriate algorithm for their applications and compare novel ideas with

existing results. The book also provides examples of clustering applications to illustrate the advantages and shortcomings of different clustering architectures and algorithms. Application areas include pattern recognition, artificial intelligence, information technology, image processing, biology, psychology, and marketing. Readers also learn how to perform cluster analysis with the C/C++ and MATLAB programming languages.

Liu has written a comprehensive text on Web mining, which consists of two parts. The first part covers the data mining and machine learning foundations, where all the essential concepts and algorithms of data mining and machine learning are presented. The second part covers the key topics of Web mining, where Web crawling, search, social network analysis, structured data extraction, information integration, opinion mining and sentiment analysis, Web usage mining, query log mining, computational advertising, and recommender systems are all treated both in breadth and in depth. His book thus brings all the related concepts and algorithms together to form an authoritative and coherent text. The book offers a rich blend of theory and practice. It is suitable for students, researchers and practitioners interested in Web mining and data mining both as a learning text and as a reference book. Professors can readily use it for classes on data mining, Web mining, and text mining. Additional teaching materials such as lecture slides, datasets, and implemented algorithms are available online.

Artificial Intelligence

Outlier Ensembles

with Practical Examples in MOA

Managing and Mining Uncertain Data

Feature Engineering for Machine Learning and Data Analytics

Mathematics for Machine Learning

Managing and Mining Uncertain Data, a survey with chapters by a variety of well known researchers in the data mining field, presents the most recent models, algorithms, and applications in the uncertain data mining field in a structured and concise way. This book is organized to make it more accessible to applications-driven practitioners for solving real problems. Also, given the lack of structurally organized information on this topic, Managing and Mining Uncertain Data provides insights which are not easily accessible elsewhere. Managing and Mining Uncertain Data is designed for a professional audience composed of researchers and practitioners in industry. This book is also suitable as a reference book for advanced-level students in computer science and engineering, as well as the ACM, IEEE, SIAM, INFORMS and AAAI Society groups.

This text takes a focused and comprehensive look at mining data represented as a graph, with the

latest findings and applications in both theory and practice provided. Even if you have minimal background in analyzing graph data, with this book you'll be able to represent data as graphs, extract patterns and concepts from the data, and apply the methodologies presented in the text to real datasets. There is a misprint with the link to the accompanying Web page for this book. For those readers who would like to experiment with the techniques found in this book or test their own ideas on graph data, the Web page for the book should be <http://www.eecs.wsu.edu/MGD>. 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 This book comprehensively covers the topic of recommender systems, which provide personalized recommendations of products or services to users based on their previous searches or purchases. Recommender system methods have been adapted to diverse applications including query log mining, social networking, news recommendations, and computational advertising. This book synthesizes both fundamental and advanced topics of a research area that has now reached maturity. The chapters of this book are organized into three categories: Algorithms and evaluation: These chapters discuss the fundamental algorithms in recommender systems, including collaborative

filtering methods, content-based methods, knowledge-based methods, ensemble-based methods, and evaluation. Recommendations in specific domains and contexts: the context of a recommendation can be viewed as important side information that affects the recommendation goals. Different types of context such as temporal data, spatial data, social data, tagging data, and trustworthiness are explored. Advanced topics and applications: Various robustness aspects of recommender systems, such as shilling systems, attack models, and their defenses are discussed. In addition, recent topics, such as learning to rank, multi-armed bandits, group systems, multi-criteria systems, and active learning systems, are introduced together with applications. Although this book primarily serves as a textbook, it will also appeal to industrial practitioners and researchers due to its focus on applications and references. Numerous examples and exercises have been provided, and a solution manual is available for instructors.

Healthcare Data Analytics

Mining Heterogeneous Information Networks

Neural Networks and Deep Learning

Recommender Systems

A Textbook

Managing and Mining Graph Data

A comprehensive overview of data mining from an algorithmic perspective, integrating related concepts from machine learning and statistics.

Privacy preserving data mining implies the "mining" of knowledge from distributed data without violating the privacy of the individual/corporations involved in contributing the data. This volume provides a comprehensive overview of available approaches, techniques and open problems in privacy preserving data mining. Crystallizing much of the underlying foundation, the book aims to inspire further research in this new and growing area.

Privacy Preserving Data Mining is intended to be accessible to industry practitioners and policy makers, to help inform future decision making and legislation, and to serve as a useful technical reference.

This book primarily discusses issues related to the mining aspects of data streams and it is unique in its primary focus on the subject. This volume covers mining aspects of data streams comprehensively: each contributed chapter contains a survey on the topic, the key

ideas in the field for that particular topic, and future research directions. The book is intended for a professional audience composed of researchers and practitioners in industry. This book is also appropriate for advanced-level students in computer science. Outlier (or anomaly) detection is a very broad field which has been studied in the context of a large number of research areas like statistics, data mining, sensor networks, environmental science, distributed systems, spatio-temporal mining, etc. Initial research in outlier detection focused on time series-based outliers (in statistics). Since then, outlier detection has been studied on a large variety of data types including high-dimensional data, uncertain data, stream data, network data, time series data, spatial data, and spatio-temporal data. While there have been many tutorials and surveys for general outlier detection, we focus on outlier detection for temporal data in this book. A large number of applications generate temporal datasets. For example, in our everyday life, various kinds of records like credit, personnel, financial, judicial, medical, etc., are all temporal. This stresses the need for an organized and detailed study of outliers with respect to such temporal data. In the past decade, there has been a lot of research on various forms of temporal data including consecutive data snapshots, series of data snapshots and data streams. Besides the initial work on time series, researchers have focused on rich forms of data including multiple data streams, spatio-temporal data, network data, community distribution data, etc. Compared to general outlier detection, techniques for temporal outlier detection are very different. In this book, we will present an organized picture of both recent and past research in temporal outlier detection. We start with the basics and then ramp up the reader to the main ideas in state-of-the-art outlier detection techniques. We motivate the importance of temporal outlier detection and brief the challenges beyond usual outlier detection. Then, we list down a taxonomy of proposed techniques for temporal outlier detection. Such techniques broadly include statistical techniques (like AR models, Markov models, histograms, neural networks), distance- and density-based approaches, grouping-based approaches (clustering, community detection), network-based approaches, and spatio-temporal outlier detection approaches. We summarize by presenting

a wide collection of applications where temporal outlier detection techniques have been applied to discover interesting outliers. Table of Contents: Preface / Acknowledgments / Figure Credits / Introduction and Challenges / Outlier Detection for Time Series and Data Sequences / Outlier Detection for Data Streams / Outlier Detection for Distributed Data Streams / Outlier Detection for Spatio-Temporal Data / Outlier Detection for Temporal Network Data / Applications of Outlier Detection for Temporal Data / Conclusions and Research Directions / Bibliography / Authors' Biographies

Statistical and Machine-Learning Data Mining:

Social Sensing

Principles and Methodologies

Machine Learning for Text

Data Streams

Mining Graph Data

Event mining encompasses techniques for automatically and efficiently extracting valuable knowledge from historical event/log data. The field, therefore, plays an important role in data-driven system management. Event Mining: Algorithms and Applications presents state-of-the-art event mining approaches and applications with a focus on computing system management. The book first explains how to transform log data in disparate formats and contents into a canonical form as well as how to optimize system monitoring. It then shows how to extract useful knowledge from data. It describes intelligent and efficient methods and algorithms to perform data-driven pattern discovery and problem determination for managing complex systems. The book also discusses data-driven approaches for the detailed diagnosis of a system issue and addresses the application of event summarization in Twitter messages (tweets).

Understanding the interdisciplinary field of event mining can be challenging as it requires familiarity with several research areas and the relevant literature is scattered in diverse publications. This book makes it easier to explore the field by providing both a good starting point for readers not familiar with the topics and a comprehensive reference for those already working in this area.

The fundamental algorithms in data mining and machine learning form the basis of data science, utilizing automated methods to analyze patterns and models for all kinds of data in applications ranging from scientific discovery to business analytics. This textbook for senior undergraduate and graduate courses provides a comprehensive, in-depth overview of data mining, machine learning and statistics, offering solid guidance for students, researchers, and practitioners. The book lays the foundations of data

analysis, pattern mining, clustering, classification and regression, with a focus on the algorithms and the underlying algebraic, geometric, and probabilistic concepts. New to this second edition is an entire part devoted to regression methods, including neural networks and deep learning.

Text analytics is a field that lies on the interface of information retrieval, machine learning, and natural language processing, and this textbook carefully covers a coherently organized framework drawn from these intersecting topics. The chapters of this textbook is organized into three categories: - Basic algorithms: Chapters 1 through 7 discuss the classical algorithms for machine learning from text such as preprocessing, similarity computation, topic modeling, matrix factorization, clustering, classification, regression, and ensemble analysis. - Domain-sensitive mining: Chapters 8 and 9 discuss the learning methods from text when combined with different domains such as multimedia and the Web. The problem of information retrieval and Web search is also discussed in the context of its relationship with ranking and machine learning methods. - Sequence-centric mining: Chapters 10 through 14 discuss various sequence-centric and natural language applications, such as feature engineering, neural language models, deep learning, text summarization, information extraction, opinion mining, text segmentation, and event detection. This textbook covers machine learning topics for text in detail. Since the coverage is extensive, multiple courses can be offered from the same book, depending on course level. Even though the presentation is text-centric, Chapters 3 to 7 cover machine learning algorithms that are often used in domains beyond text data. Therefore, the book can be used to offer courses not just in text analytics but also from the broader perspective of machine learning (with text as a backdrop). This textbook targets graduate students in computer science, as well as researchers, professors, and industrial practitioners working in these related fields. This textbook is accompanied with a solution manual for classroom teaching.

This book discusses a variety of methods for outlier ensembles and organizes them by the specific principles with which accuracy improvements are achieved. In addition, it covers the techniques with which such methods can be made more effective. A formal classification of these methods is provided, and the circumstances in which they work well are examined. The authors cover how outlier ensembles relate (both theoretically and practically) to the ensemble techniques used commonly for other data mining problems like classification. The similarities and (subtle) differences in the ensemble techniques for the classification and outlier detection problems are explored. These subtle differences do impact the design of ensemble algorithms for the latter problem. This book can be used for courses in data mining and related curricula. Many illustrative examples and exercises are provided in order to facilitate classroom teaching. A familiarity is assumed to the outlier detection problem and also to generic problem of ensemble analysis in classification. This is because many of the ensemble methods discussed in this book are adaptations from their counterparts in the classification domain. Some techniques

explained in this book, such as wagging, randomized feature weighting, and geometric subsampling, provide new insights that are not available elsewhere. Also included is an analysis of the performance of various types of base detectors and their relative effectiveness. The book is valuable for researchers and practitioners for leveraging ensemble methods into optimal algorithmic design.

Data Clustering

Semantic Data Mining

Privacy Preserving Data Mining

Techniques for Better Predictive Modeling and Analysis of Big Data, Third Edition

Social Network Data Analytics

Outlier Detection for Temporal Data

Advances in hardware technology have increased the capability to store and record personal data. This has caused concerns that personal data may be abused. This book proposes a number of techniques to perform the data mining tasks in a privacy-preserving way. This edited volume contains surveys by distinguished researchers in the privacy field. Each survey includes the key research content as well as future research directions of a particular topic in privacy. The book is designed for researchers, professors, and advanced-level students in computer science, but is also suitable for practitioners in industry.

This two-volume handbook presents a collection of novel methodologies with applications and illustrative examples in the areas of data-driven computational social sciences. Throughout this handbook, the focus is kept specifically on business and consumer-oriented applications with interesting sections ranging from clustering and network analysis, meta-analytics, memetic algorithms, machine learning, recommender systems methodologies, parallel pattern mining and data mining to specific applications in market segmentation, travel, fashion or entertainment analytics. A must-read for anyone in data-analytics, marketing, behavior modelling and computational social science, interested in the latest applications of new computer science methodologies. The chapters are contributed by leading experts in the associated fields. The chapters cover technical aspects at different levels, some of which are introductory and could be used for teaching. Some chapters aim at building a common understanding of the methodologies and recent application areas including the introduction of new theoretical results in the complexity of core problems. Business and marketing professionals may use the book to familiarize themselves with some important foundations of data science. The work is a good starting point to establish an open dialogue of communication between professionals and researchers from different fields. Together, the two volumes present a number of different new directions in Business and Customer Analytics with an emphasis in personalization of services, the development of new mathematical models and new algorithms, heuristics and metaheuristics applied to the challenging problems in the field. Sections of the book have introductory material to more specific and advanced themes in some of the chapters, allowing the volumes to be used as an advanced textbook. Clustering, Proximity Graphs, Pattern Mining, Frequent Itemset Mining, Feature Engineering, Network and Community Detection, Network-based Recommending Systems and Visualization, are some of the topics in the first volume. Techniques on Memetic Algorithms and their applications to Business Analytics and Data Science are surveyed in the second volume; applications in Team Orienteering, Competitive Facility-location, and Visualization of Products and Consumers are also discussed. The second volume also includes an introduction to Meta-Analytics, and to the application areas of Fashion and Travel Analytics. Overall, the two-volume set helps to describe

some fundamentals, acts as a bridge between different disciplines, and presents important results in a rapidly moving field combining powerful optimization techniques allied to new mathematical models critical for personalization of services. Academics and professionals working in the area of business analytics, data science, operations research and marketing will find this handbook valuable as a reference. Students studying these fields will find this handbook useful and helpful as a secondary textbook.

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.

Advances in hardware technology have led to an ability to collect data with the use of a variety of sensor technologies. In particular sensor nodes have become cheaper and more efficient, and have even been integrated into day-to-day devices of use, such as mobile phones. This has led to a much larger scale of applicability and mining of sensor data sets. The human-centric aspect of sensor data has created tremendous opportunities in integrating social aspects of sensor data collection into the mining process. Managing and Mining Sensor Data is a contributed volume by prominent leaders in this field, targeting advanced-level students in computer science as a secondary text book or reference. Practitioners and researchers working in this field will also find this book useful.

Algorithms and Applications

Exploring Hyperlinks, Contents, and Usage Data

Privacy-Preserving Data Mining

Outlier Analysis

An Introduction

Data Mining and Analysis

Ontologies are now increasingly used to integrate, and organize data and knowledge, particularly in data and knowledge-intensive applications in both research and industry. The book is devoted to semantic data mining – a data mining approach where domain ontologies are used as background knowledge, and where the new challenge is to mine knowledge encoded in domain ontologies and knowledge graphs, rather than only purely empirical data. The introductory chapters of the book provide theoretical foundations of both data mining and ontology representation. Taking a unified perspective, the book then covers several methods for semantic data mining, addressing tasks such as pattern mining, classification and similarity-based approaches. It attempts to provide state-of-the-art answers to specific challenges and peculiarities of data mining with use of ontologies, in particular: How to deal with incompleteness of knowledge and the so-called Open World Assumption? What is a truly “semantic” similarity measure? The book contains several chapters with examples of applications of semantic data mining. The examples start from a scenario with moderate use of lightweight ontologies for knowledge graph enrichment and end with a full-fledged scenario of an intelligent knowledge discovery assistant using complex domain ontologies for meta-mining, i.e., an ontology-based meta-

learning approach to full data mining processes. The book is intended for researchers in the fields of semantic technologies, knowledge engineering, data science, and data mining, and developers of knowledge-based systems and applications. A hands-on approach to tasks and techniques in data stream mining and real-time analytics, with examples in MOA, a popular freely available open-source software framework. Today many information sources—including sensor networks, financial markets, social networks, and healthcare monitoring—are so-called data streams, arriving sequentially and at high speed. Analysis must take place in real time, with partial data and without the capacity to store the entire data set. This book presents algorithms and techniques used in data stream mining and real-time analytics. Taking a hands-on approach, the book demonstrates the techniques using MOA (Massive Online Analysis), a popular, freely available open-source software framework, allowing readers to try out the techniques after reading the explanations. The book first offers a brief introduction to the topic, covering big data mining, basic methodologies for mining data streams, and a simple example of MOA. More detailed discussions follow, with chapters on sketching techniques, change, classification, ensemble methods, regression, clustering, and frequent pattern mining. Most of these chapters include exercises, an MOA-based lab session, or both. Finally, the book discusses the MOA software, covering the MOA graphical user interface, the command line, use of its API, and the development of new methods within MOA. The book will be an essential reference for readers who want to use data stream mining as a tool, researchers in innovation or data stream mining, and programmers who want to create new algorithms for MOA.

Social network analysis applications have experienced tremendous advances within the last few years due in part to increasing trends towards users interacting with each other on the internet. Social networks are organized as graphs, and the data on social networks takes on the form of massive streams, which are mined for a variety of purposes. Social Network Data Analytics covers an important niche in the social network analytics field. This edited volume, contributed by prominent researchers in this field, presents a wide selection of topics on social network data mining such as Structural Properties of Social Networks, Algorithms for Structural Discovery of Social Networks and Content Analysis in Social Networks. This book is also unique in focussing on the data analytical aspects of social networks in the internet scenario, rather than the traditional sociology-driven emphasis prevalent in the existing books, which do not focus on the unique data-intensive characteristics of online social networks. Emphasis is placed on simplifying the content so that students and practitioners benefit from this book. This book targets advanced level students and researchers concentrating on computer science as a secondary text or reference book. Data mining, database, information security, electronic commerce and machine learning professionals will find this book a valuable asset, as well as primary associations such as ACM, IEEE and Management Science.

Real-world physical and abstract data objects are interconnected, forming gigantic, interconnected networks. By structuring these data objects and interactions between these objects into multiple types, such networks become semi-structured heterogeneous information networks. Most real-world applications that handle big data, including interconnected social media and social networks, scientific, engineering, or medical information systems, online e-commerce systems, and most database systems, can be structured into heterogeneous information networks. Therefore, effective analysis of large-scale heterogeneous information networks poses an interesting but critical challenge. In this book, we investigate the principles and methodologies of mining heterogeneous information networks. Departing from many existing network models that view interconnected data as

homogeneous graphs or networks, our semi-structured heterogeneous information network model leverages the rich semantics of typed nodes and links in a network and uncovers surprisingly rich knowledge from the network. This semi-structured heterogeneous network modeling leads to a series of new principles and powerful methodologies for mining interconnected data, including: (1) rank-based clustering and classification; (2) meta-path-based similarity search and mining; (3) relation strength-aware mining, and many other potential developments. This book introduces this new research frontier and points out some promising research directions. Table of Contents: Introduction / Ranking-Based Clustering / Classification of Heterogeneous Information Networks / Meta-Path-Based Similarity Search / Meta-Path-Based Relationship Prediction / Relation Strength-Aware Clustering with Incomplete Attributes / User-Guided Clustering via Meta-Path Selection / Research Frontiers

Frequent Pattern Mining

Models and Algorithms

Clustering for Data Mining

Mining Text Data

Frequent Pattern Mining in Health Care Database

Data Mining and Machine Learning

This textbook explores the different aspects of data mining from the fundamentals to the complex data types and their applications, capturing the wide diversity of problem domains for data mining issues. It goes beyond the traditional focus on data mining problems to introduce advanced data types such as text, time series, discrete sequences, spatial data, graph data, and social networks. Until now, no single book has addressed all these topics in a comprehensive and integrated way. The chapters of this book fall into one of three categories: Fundamental chapters: Data mining has four main problems, which correspond to clustering, classification, association pattern mining, and outlier analysis. These chapters comprehensively discuss a wide variety of methods for these problems. Domain chapters: These chapters discuss the specific methods used for different domains of data such as text data, time-series data, sequence data, graph data, and spatial data. Application chapters: These chapters study important applications such as stream mining, Web mining, ranking, recommendations, social networks, and privacy preservation. The domain chapters also have an applied flavor. Appropriate for both introductory and advanced data mining courses, Data Mining: The Textbook balances mathematical details and intuition. It contains the necessary mathematical details for professors and researchers, but it is presented in a simple and intuitive style to improve accessibility for students and industrial practitioners (including those with a limited mathematical background). Numerous illustrations, examples, and exercises are included, with an emphasis on semantically interpretable examples. Praise for Data Mining: The Textbook - "As I read through this book, I have already decided to use it in my classes. This is a book written by an outstanding researcher who has made fundamental contributions to data mining, in a way that is both accessible and up to date. The book is complete with theory and practical use cases. It's a must-have for students and professors alike!" -- Qiang Yang, Chair of Computer Science and Engineering at Hong Kong University of Science and Technology "This is the most amazing and comprehensive text book on data mining. It covers not

only the fundamental problems, such as clustering, classification, outliers and frequent patterns, and different data types, including text, time series, sequences, spatial data and graphs, but also various applications, such as recommenders, Web, social network and privacy. It is a great book for graduate students and researchers as well as practitioners." -- Philip S. Yu, UIC Distinguished Professor and Wexler Chair in Information Technology at University of Illinois at Chicago

Integrates social media, social network analysis, and data mining to provide an understanding of the potentials of social media mining.

Often considered more as an art than a science, the field of clustering has been dominated by learning through examples and by techniques chosen almost through trial-and-error. Even the most popular clustering methods--K-Means for partitioning the data set and Ward's method for hierarchical clustering--have lacked the theoretical attention that wou

Feature engineering plays a vital role in big data analytics. Machine learning and data mining algorithms cannot work without data. Little can be achieved if there are few features to represent the underlying data objects, and the quality of results of those algorithms largely depends on the quality of the available features. Feature Engineering for Machine Learning and Data Analytics provides a comprehensive introduction to feature engineering, including feature generation, feature extraction, feature transformation, feature selection, and feature analysis and evaluation. The book presents key concepts, methods, examples, and applications, as well as chapters on feature engineering for major data types such as texts, images, sequences, time series, graphs, streaming data, software engineering data, Twitter data, and social media data. It also contains generic feature generation approaches, as well as methods for generating tried-and-tested, hand-crafted, domain-specific features. The first chapter defines the concepts of features and feature engineering, offers an overview of the book, and provides pointers to topics not covered in this book. The next six chapters are devoted to feature engineering, including feature generation for specific data types. The subsequent four chapters cover generic approaches for feature engineering, namely feature selection, feature transformation based feature engineering, deep learning based feature engineering, and pattern based feature generation and engineering. The last three chapters discuss feature engineering for social bot detection, software management, and Twitter-based applications respectively. This book can be used as a reference for data analysts, big data scientists, data preprocessing workers, project managers, project developers, prediction modelers, professors, researchers, graduate students, and upper level undergraduate students. It can also be used as the primary text for courses on feature engineering, or as a supplement for courses on machine learning, data mining, and big data analytics.

Building Reliable Systems on Unreliable Data

Machine Learning

Introduction to Data Mining

Business and Consumer Analytics: New Ideas

Fundamental Concepts and Algorithms

Event Mining

This book provides comprehensive coverage of the field of outlier analysis from a computer science point of view. It integrates methods from data mining, machine learning, and statistics within the computational framework and therefore appeals to multiple communities. The chapters of this book can be organized into three categories: Basic algorithms: Chapters 1 through 7 discuss the fundamental algorithms for outlier analysis, including probabilistic and statistical methods, linear methods, proximity-based methods, high-dimensional (subspace) methods, ensemble methods, and supervised methods. Domain-specific methods: Chapters 8 through 12 discuss outlier detection algorithms for various domains of data, such as text, categorical data, time-series data, discrete sequence data, spatial data, and network data. Applications: Chapter 13 is devoted to various applications of outlier analysis. Some guidance is also provided for the practitioner. The second edition of this book is more detailed and is written to appeal to both researchers and practitioners. Significant new material has been added on topics such as kernel methods, one-class support-vector machines, matrix factorization, neural networks, outlier ensembles, time-series methods, and subspace methods. It is written as a textbook and can be used for classroom teaching.

At the intersection of computer science and healthcare, data analytics has emerged as a promising tool for solving problems across many healthcare-related disciplines. Supplying a comprehensive overview of recent healthcare analytics research, Healthcare Data Analytics provides a clear understanding of the analytical techniques currently available to solve healthcare problems. The book details novel techniques for acquiring, handling, retrieving, and making best use of healthcare data. It analyzes recent developments in healthcare computing and discusses emerging technologies that can help improve the health and well-being of patients. Written by prominent researchers and experts working in the healthcare domain, the book sheds light on many of the computational challenges in the field of medical informatics. Each chapter in the book is structured as a "survey-style" article discussing the prominent research issues and the advances made on that research topic. The book is divided into three major categories: Healthcare Data Sources and Basic Analytics - details the various healthcare data sources and analytical techniques used in the processing and analysis of such data Advanced Data Analytics for Healthcare - covers advanced analytical methods, including clinical prediction models, temporal pattern mining methods, and visual analytics Applications and Practical Systems for Healthcare - covers the applications of data analytics to pervasive healthcare, fraud detection, and drug discovery along with systems for medical imaging and decision support Computer scientists are usually not trained in domain-specific medical concepts, whereas medical practitioners and researchers have limited exposure to the data analytics area. The contents of this book will help to bring together these diverse communities by

carefully and comprehensively discussing the most relevant contributions from each domain.