Data Mining Techniques In Parallel And Distributed

JULISSA KIRSTEN

Parallel Computing: Fundamentals, Applications and New Directions BoD – Books on Demand
With the unprecedented growth-rate at which data is being collected and stored electronically today in almost all fields of human endeavor, the efficient extraction of useful information from the data available is becoming an increasing scientific challenge and a massive economic need. This book presents thoroughly reviewed and revised full versions of papers presented at a workshop on the topic held during KDD'99 in San Diego, California, USA in August 1999 complemented by several invited chapters and a detailed introductory survey in order to provide complete coverage of the relevant issues. The contributions presented cover all major tasks in data mining including parallel and distributed mining frameworks, associations, sequences, clustering, and classification. All in all, the volume presents the state of the art in the young and dynamic field of parallel and distributed data mining methods. It will be a valuable source of reference for researchers and professionals.
Data Mining and Knowledge Discovery in Real Life Applications John Wiley & Sons
The refereed proceedings of the International Symposium on Parallel and Distributed Processing and Applications, ISPA 2003, held in Aizu, Japan in July 2003. The 30 revised full papers and 9 revised short papers presented together with abstracts of 4 keynotes were carefully reviewed and selected from numerous submissions. The papers are organized in topical sections on applications on Web-based and intranet systems, compiler and optimization techniques, network routing, performance evaluation of parallel systems, wireless communication and mobile computing, parallel topology, data mining and evolutionary computing, image processing and modeling, network security, and database and multimedia systems.
Data Mining Springer Science & Business Media
The book presents the latest, high-quality, technical contributions and research findings in the areas of data management and smart computing, big data management, artificial intelligence and data analytics, along with advances in

network technologies. It discusses state-of-the-art topics as well as the challenges and solutions for future development. It includes original and previously unpublished international research work highlighting research domains from different perspectives. This book is mainly intended for researchers and practitioners in academia and industry.

Parallel Data Mining for Association Rules on Shared-memory Multi-processors Springer Science & Business Media

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

Advanced Data Mining Techniques SK Research Group of Companies

High Performance Data Mining: Scaling Algorithms, Applications and Systems brings together in one place important contributions and up-to-date research results in this fast moving area. High Performance Data Mining: Scaling Algorithms, Applications and Systems serves as an excellent reference, providing insight into some of the most challenging research issues in the field.

Data Mining for Association Rules and Sequential Patterns Springer Science & Business Media

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

Data Mining Springer Science & Business Media

Dr.P.Alagesh Kannan, Assistant Professor, Department of Computer Science, Madurai Kamaraj University College, Madurai,Tamil Nadu, India. Dr.J.Saravanesh, Assistant Professor, Department of Computer Science, Madurai Kamaraj University College, Madurai,Tamil Nadu, India.

Data Mining Cambridge University Press

Based around eleven international real life case studies and including contributions from leading experts in the field this groundbreaking book explores the need for the grid-enabling of data mining applications and provides a comprehensive study of the technology, techniques and management skills necessary to create them. This book provides a simultaneous design blueprint, user guide, and research agenda for current and future developments and will appeal to a broad audience; from developers and users of data mining and grid technology, to advanced undergraduate and postgraduate students interested in this field.

New Fundamental Technologies in Data Mining BoD – Books on Demand

Data Analysis in the Cloud introduces and discusses models, methods, techniques, and systems to analyze the large number of digital data sources available on the Internet using the computing and storage facilities of the cloud. Coverage includes scalable data mining and knowledge discovery techniques together with cloud computing concepts, models, and systems. Specific sections focus on map-reduce and NoSQL models. The book also includes techniques for conducting high-performance distributed analysis of large data on clouds. Finally, the book examines research trends such as Big Data pervasive computing, data-intensive exascale computing, and massive social network analysis. Introduces data analysis techniques and cloud computing concepts Describes cloud-based models and systems for Big Data analytics Provides examples of the state-of-the-art in cloud data analysis Explains how to develop large-scale data mining applications on clouds Outlines the main research trends in the area of scalable Big Data analysis

Parallel Data Mining Revisited Elsevier

Knowledge Discovery and Measures of Interest is a reference book for knowledge discovery researchers, practitioners, and students. The knowledge discovery researcher will find that the material provides a theoretical foundation for measures of interest in data mining applications where diversity measures are used to rank summaries generated from databases. The knowledge discovery practitioner will find solid empirical evidence on which to base decisions regarding the choice of measures in data mining applications. The knowledge discovery student in a senior undergraduate or graduate course in databases and data mining will find the book is a good introduction to the concepts and techniques of measures of interest. In Knowledge Discovery and Measures of Interest, we study two closely related steps in any knowledge discovery system: the generation of discovered knowledge; and the interpretation and evaluation of discovered knowledge. In the generation step, we study data summarization, where a single dataset can be generalized in many different ways and to many different levels of granularity according to domain generalization graphs. In the interpretation and evaluation step, we study diversity measures as heuristics for ranking the interestingness of the summaries generated. The objective of this work is to introduce and evaluate a technique for ranking the interestingness of discovered patterns in data. It consists of four primary goals: To introduce domain generalization graphs for describing and guiding the generation of summaries from databases. To introduce and evaluate serial and parallel algorithms that traverse the domain generalization space described by the domain generalization graphs. To introduce and evaluate diversity measures as heuristic measures of interestingness for ranking summaries generated from databases. To develop the preliminary foundation for a theory of interestingness within the context of ranking summaries generated from databases. Knowledge Discovery and Measures of Interest is suitable as a secondary text in a graduate level course and as a reference for researchers and practitioners in industry.

Applied Parallel Computing: Advanced Scientific Computing John Wiley & Sons
This integrated collection covers a range of parallelization platforms, concurrent programming frameworks and machine learning settings, with case studies.
Data Mining Techniques in Grid Computing Environments BoD – Books on Demand
The 15th Workshop on Languages and Compilers for Parallel Computing was held in July 2002 at the University of Maryland, College Park. It was jointly sponsored by the Department of Computer Science at the University of Ma- land and the University of Maryland Institute for Advanced Computer Studies (UMIACS).LCPC2002broughttogetherover60researchersfromacademiaand research institutions from many countries. The program of 26 papers was selected from 32 submissions. Each paper was reviewed by at least three Program Committee members and sometimes by additional reviewers. Prior to the workshop, revised versions of accepted papers were informally published on the workshop's website and in a paper proceedings that was distributed at the meeting. This year, the workshopwas organizedinto sessions of papers on related topics, and each session consisted of two to three 30-minute presentations.Based on feedback from the workshop,the papers were revised and submitted for inclusion in the formal proceedings published in this volume. Two papers were presented at the workshop but later withdrawn from the ?nal proceedings by their authors. We were very lucky to have Bill Carlson from the Department of Defense give the LCPC 2002 keynote speech on "UPC: A C Language for Shared M- ory Parallel Programming." Bill gave an excellent overview of the features and programming model of the UPC parallel programming language.
Advanced Data Mining Techniques Springer
This book on data mining explores a broad set of ideas and presents some of the state-of-the-art research in this field. The book is triggered by pervasive applications that retrieve knowledge from real-world big data. Data mining finds applications in the entire spectrum of science and technology including basic sciences to life sciences and medicine, to social, economic, and cognitive sciences, to engineering and computers. The chapters discuss various applications and research frontiers in data mining with algorithms and implementation details for use in real-world. This can be through characterization, classification, discrimination, anomaly detection, association, clustering, trend or evolution prediction, etc. The intended audience of this book will mainly consist of researchers, research students, practitioners, data analysts, and business professionals who seek information on the various data mining techniques and their applications.
Languages and Compilers for Parallel Computing Springer Science & Business Media
This volume is aiming at a wide range of readers and researchers in the area of Big Data by presenting the recent advances in the fields of Big Data Analysis, as well as the techniques and tools used to analyze it. The book includes 10 distinct chapters providing a concise introduction to Big Data Analysis and recent Techniques and Environments for Big Data Analysis. It gives insight into how the expensive fitness evaluation of evolutionary learning can play a vital role in big data analysis by adopting Parallel, Grid, and Cloud computing environments.
Data Mining for Scientific and Engineering Applications Springer
With big data analytics comes big insights into profitability Big data is big business. But having the data and the computational power to process it isn't nearly enough to produce meaningful results. Big Data, Data Mining, and Machine Learning: Value Creation for Business Leaders and Practitioners is a complete resource for technology and marketing executives looking to cut through the hype and produce real results that hit the bottom line. Providing an engaging, thorough overview of the current state of big data analytics and the growing trend toward high performance computing architectures, the book is a detail-driven look into how big data analytics can be leveraged to foster positive change and drive efficiency. With continued exponential growth in data and ever more competitive markets, businesses must adapt quickly to gain every competitive advantage available. Big data analytics can serve as the linchpin for initiatives that drive business, but only if the underlying technology and analysis is fully understood and appreciated by engaged stakeholders. This book provides a view into the topic that executives, managers, and practitioners require, and includes: A complete overview of big data and its notable characteristics Details on high performance computing architectures for analytics, massively parallel processing (MPP), and in-memory databases Comprehensive coverage of data mining, text analytics, and machine learning algorithms A discussion of explanatory and predictive modeling, and how they can be applied to decision-making processes Big Data, Data Mining, and Machine Learning provides technology and marketing executives with the complete resource that has been notably absent from the veritable libraries of published books on the topic. Take control of your organization's big data analytics to produce real results with a resource that is comprehensive in scope and light on hyperbole.
Parallel Data Mining with the Message Passing Interface Standard on Clusters of Personal Computers World Scientific
Owing to continuous advances in the computational power of handheld devices like smartphones and tablet computers, it has become possible to perform Big Data operations including modern data mining processes onboard these small devices. A decade of research has proved the feasibility of what has been termed as Mobile Data Mining, with a focus on one mobile device running data mining processes. However, it is not before 2010 until the authors of this book initiated the Pocket Data Mining (PDM) project exploiting the seamless communication among handheld devices performing data analysis tasks that were infeasible until recently. PDM is the process of collaboratively extracting knowledge from distributed data streams in a mobile computing environment. This book provides the reader with an in-depth treatment on this emerging area of research. Details of techniques used and thorough experimental studies are given. More importantly and exclusive to this book, the authors provide detailed practical guide on the deployment of PDM in the mobile environment. An important extension to the basic implementation of PDM dealing with concept drift is also reported. In the era of Big Data, potential applications of paramount importance offered by PDM in a variety of domains including security, business and telemedicine are discussed.
Knowledge Discovery and Measures of Interest Springer
Piles of personal computers (PoPCs) have begun to challenge the performance of the traditional Massively Parallel Processors (MPPs) and the less traditional networks of workstations (NOWs) as platforms for parallel computing. Large clusters of PCs have reached and at times exceeded the performance of modern MPPs at a fraction of the cost. Built with commodity components, these clusters can be constructed for about half the cost of a comparable NOW. The primary competing operating systems (OIS) in use on PoPCs

are Linux and Windows NT. This thesis investigation compares the performance of an NT cluster with that of a Linux cluster, a NOW, and an MPP. A comparison of the MPI tools available for NT is also accomplished. These comparisons are made using the Pallas benchmark suite for MPI and a parallel data mining algorithm. This data mining technique, known as the Genetic Rule and Classifier Construction Environment (GRaCCE), uses a genetic algorithm to mine decision rules from data. Results from experimentation and statistical analysis have produced three important conclusions. First, NT clusters are viable, cost effective alternatives to Linux clusters, NOWs, and MPPs for parallel computing. Second, the two primary communication libraries currently available for NT-PaTENT MPI and MPI/Pro- are statistically equivalent in performance. Third, the parallel GRaCCE algorithm is capable of relatively good speedup and efficiency, even for significantly unbalanced processor workloads, if the effects of first loop iteration caching are ignored.

Data Management, Analytics and Innovation Springer Science & Business Media

This book presents four different ways of theoretical and practical advances and applications of data mining in different promising areas like Industrialist, Biological, and Social. Twenty six chapters cover different special topics with proposed novel ideas. Each chapter gives an overview of the subjects and some of the chapters have cases with offered data mining solutions. We hope that this book will be a useful aid in showing a right way for the students, researchers and practitioners in their studies.

Large-Scale Parallel Data Mining CRC Press

The continual explosion of information technology and the need for better data collection and management methods has made data mining an even more relevant topic of study. Books on data mining tend to be either broad and introductory or focus on some very specific technical aspect of the field. This book is a series of seventeen edited OC student-authored lecturesOCO which explore in depth the core of data mining (classification, clustering and association rules) by offering overviews that include both analysis and insight. The initial chapters lay a framework of data mining techniques by explaining some of the basics such as applications of Bayes Theorem, similarity measures, and decision trees. Before focusing on the pillars of classification, clustering and association rules, the book also considers alternative candidates such as point estimation and genetic algorithms. The book''s discussion of classification includes an introduction to decision tree algorithms, rule-based algorithms (a popular alternative to decision trees) and distance-based algorithms. Five of the lecture-chapters are devoted to the concept of clustering or unsupervised classification. The functionality of hierarchical and partitional clustering algorithms is also covered as well as the efficient and scalable clustering algorithms used in large databases. The concept of association rules in terms of basic algorithms, parallel and distributive algorithms and advanced measures that help determine the value of association rules are discussed. The final chapter discusses algorithms for spatial data mining. Sample Chapter(s). Chapter 1: Point Estimation Algorithms (397 KB). Contents: Point Estimation Algorithms; Applications of Bayes Theorem; Similarity Measures; Decision Trees; Genetic Algorithms; Classification: Distance Based Algorithms; Decision Tree-Based Algorithms; Covering (Rule-Based) Algorithms; Clustering: An Overview; Clustering Hierarchical Algorithms; Clustering Partitional Algorithms; Clustering: Large Databases; Clustering Categorical Attributes; Association Rules: An Overview; Association Rules: Parallel and Distributed Algorithms; Association Rules: Advanced Techniques and Measures; Spatial Mining: Techniques and Algorithms. Readership: An introductory data mining textbook or a technical data mining book for an upper level undergraduate or graduate level course."

Association Rule Hiding for Data Mining Springer Science & Business Media

Advances in Data Mining Knowledge Discovery and Applications aims to help data miners, researchers, scholars, and PhD students who wish to apply data mining techniques. The primary contribution of this book is highlighting frontier fields and implementations of the knowledge discovery and data mining. It seems to be same things are repeated again. But in general, same approach and techniques may help us in different fields and expertise areas. This book presents knowledge discovery and data mining applications in two different sections. As known that, data mining covers areas of statistics, machine learning, data management and databases, pattern recognition, artificial intelligence, and other areas. In this book, most of the areas are covered with different data mining applications. The eighteen chapters have been classified in two parts: Knowledge Discovery and Data Mining Applications.