

Discovering Knowledge In Data An Introduction To Data Mining 2nd Edition

This book contains thoroughly refereed extended papers from the Second International Workshop on Knowledge Discovery from Sensor Data, Sensor-KDD 2008, held in Las Vegas, NV, USA, in August 2008. The 12 revised papers presented together with an invited paper were carefully reviewed and selected from numerous submissions. The papers feature important aspects of knowledge discovery from sensor data, e.g., data mining for diagnostic debugging; incremental histogram distribution for change detection; situation-aware adaptive visualization; WiFi mining; mobile sensor data mining; incremental anomaly detection; and spatiotemporal neighborhood discovery for sensor data.

Currently there are major challenges in data mining applications in the geosciences. This is due primarily to the fact that there is a wealth of available mining data amid an absence of the knowledge and expertise necessary to analyze and accurately interpret the same data. Most geoscientists have no practical knowledge or experience using data mining techniques. For the few that do, they typically lack expertise in using data mining software and in selecting the most appropriate algorithms for a given application. This leads to a paradoxical scenario of "rich data but poor knowledge". The true solution is to apply data mining techniques in geosciences databases and to modify these techniques for practical applications. Authored by a global thought leader in data mining, Data Mining and Knowledge Discovery for Geoscientists addresses these challenges by summarizing the latest developments in geosciences data mining and arming scientists with the ability to apply key concepts to effectively analyze and interpret vast amounts of critical information. Focuses on 22 of data mining's most practical algorithms and popular application samples Features 36 case studies and end-of-chapter exercises unique to the geosciences to underscore key data mining applications Presents a practical and integrated system of data mining and knowledge discovery for geoscientists Rigorous yet broadly accessible to geoscientists, engineers, researchers and programmers in data mining Introduces widely used algorithms, their basic principles and conditions of applications, diverse case studies, and suggests algorithms that may be suitable for specific applications

Solutions Manual to accompany Statistical Data Analytics: Foundations for Data Mining, Informatics, and Knowledge Discovery A comprehensive introduction to statistical methods for data mining and knowledge discovery. Extensive solutions using actual data (with sample R programming code) are provided, illustrating diverse informatic sources in genomics, biomedicine, ecological remote sensing, astronomy, socioeconomics, marketing, advertising and finance, among many others.

This book constitutes the thoroughly refereed post-workshop proceedings at PAKDD Workshops 2017, held in conjunction with PAKDD, the 21st Pacific-Asia Conference on Knowledge Discovery and Data Mining in May 2017 in Jeju, South Korea. The 17 revised papers presented were carefully reviewed and selected from 38 submissions. The workshops affiliated with PAKDD 2017 include: Workshop on Machine Learning for Sensory Data Analysis (MLSDA), Workshop on Biologically Inspired Data Mining Techniques (BDM), Pacific Asia Workshop on Intelligence and Security Informatics (PAISI), and Workshop on Data Mining in Business Process Management (DM-BPM).

Multimedia Data Mining and Knowledge Discovery

Discovering Knowledge in Data

PAKDD 2017 Workshops, MLSDA, BDM, DM-BPM Jeju, South Korea, May 23, 2017, Revised Selected Papers

Advances in Knowledge Discovery and Data Mining

Challenges and Realities

Discovering Knowledge from Hypertext Data

The definitive book on mining the Web from the preeminent authority.

This open access book is the first to systematically introduce the principles of urban informatics and its application to every aspect of the city that involves its functioning, control, management, and future planning. It introduces new models and tools being developed to understand and implement these technologies that enable cities to function more efficiently to become smart and sustainable. The smart city has quickly emerged as computers have become ever smaller to the point where they can be embedded into the very fabric of the city, as well as being central to new ways in which the population can communicate and act. When cities are wired in this way, they have the potential to become sentient and responsive, generating massive streams of big data in real time as well as providing immense opportunities for extracting new forms of urban data through crowdsourcing. This book offers a comprehensive review of the methods that form the core of urban informatics from various kinds of urban remote sensing to new approaches to machine learning and statistical modelling. It provides a detailed technical introduction to the wide array of tools information scientists need to develop the key urban analytics that are fundamental to learning about the smart city, and it outlines ways in which these tools can be used to inform design and policy so that cities can become more efficient with a greater concern for environment and equity.

Covering research at the frontier of this field, Privacy-Aware Knowledge Discovery: Novel Applications and New Techniques presents state-of-the-art privacy-preserving data mining techniques for application domains, such as medicine and social networks, that face the increasing heterogeneity and complexity of new forms of data. Renowned authorities from prominent organizations not only cover well-established results they also explore complex domains where privacy issues are generally clear and well defined, but the solutions are still preliminary and in continuous development. Divided into seven parts, the book provides in-depth coverage of the most novel reference scenarios for privacy-preserving techniques. The first part gives general techniques that can be applied to various applications discussed in the rest of the book. The second section focuses on the sanitization of network traces and privacy in data stream mining. After the third part on privacy in spatio-temporal data mining and mobility data analysis, the book examines time series analysis in the fourth section, explaining how a perturbation method and a segment-based method can tackle privacy issues of time series data. The fifth section on biomedical data addresses genomic data as well as the problem of privacy-aware information sharing of health data. In the sixth section on web applications, the book deals with query log mining and web recommender systems. The final part on social networks analyzes privacy issues related to the management of social network data under different perspectives. While several new results have recently occurred in the privacy, database, and data mining research communities, a uniform presentation of up-to-date techniques and applications is lacking. Filling this void, Privacy-Aware Knowledge Discovery presents novel algorithms, patterns, and models, along with a significant collection of open problems for future investigation.

Eight sections of this book span fundamental issues of knowledge discovery, classification and clustering, trend and deviation analysis, dependency derivation, integrated discovery systems, augmented database systems and application case studies. The appendices provide a list of terms used in the literature of the field of data mining and knowledge discovery in databases, and a list of online resources for the KDD researcher.

Data Mining

Medical Data Mining and Knowledge Discovery

Second International Workshop, Sensor-KDD 2008, Las Vegas, NV, USA, August 24-27, 2008, Revised Selected Papers

A Knowledge Discovery Approach

The Art of Excavating Data for Knowledge Discovery

This book presents a specific and unified approach to Knowledge Discovery and Data Mining, termed IFN for Information Fuzzy Network methodology. Data Mining (DM) is the science of modelling and generalizing common patterns from large sets of multi-type data. DM is a part of KDD, which is the overall process for Knowledge Discovery in Databases. The accessibility and abundance of information today makes this a topic of particular importance and need. The book has three main parts complemented by appendices as well as software and project data that are accessible from the book's web site (<http://www.eng.tau.ac.il/~maimonlifn-kdgm/>). Part I (Chapters 1-4) starts with the topic of KDD and DM in general and makes reference to other works in the field, especially those related to the information theoretic approach. The remainder of the book presents our work, starting with the IFN theory and algorithms. Part II (Chapters 5-6) discusses the methodology of application and includes case studies. Then in Part III (Chapters 7-9) a comparative study is presented, concluding with some advanced methods and open problems. The IFN, being a generic methodology, applies to a variety of fields, such as manufacturing, finance, health care, medicine, insurance, and human resources. The appendices expand on the relevant theoretical background and present descriptions of sample projects (including detailed results).

This text surveys research from the fields of data mining and information visualisation and presents a case for techniques by which information visualisation can be used to uncover real knowledge hidden away in large databases.

Data Mining is the science and technology of exploring data in order to discover previously unknown patterns. It is a part of the overall process of Knowledge Discovery in Databases (KDD). The accessibility and abundance of information today makes data mining a matter of considerable importance and necessity. This book provides an introduction to the field with an emphasis on advanced decomposition methods in general data mining tasks and for classification tasks in particular. The book presents a complete methodology for decomposing classification problems into smaller and more manageable sub-problems that are solvable by using existing tools. The various elements are then joined together to solve the initial problem. The benefits of decomposition methodology in data mining include: increased performance (classification accuracy); conceptual simplification of the problem; enhanced feasibility for huge databases; clearer and more comprehensible results; reduced runtime by solving smaller problems and by using parallel/distributed computation; and the opportunity of using different techniques for individual sub-problems.

This book integrates two areas of computer science, namely data mining and evolutionary algorithms. Both these areas have become increasingly popular in the last few years, and their integration is currently an active research area. In general, data mining consists of extracting knowledge from data. The motivation for applying evolutionary algorithms to data mining is that evolutionary algorithms are robust search methods which perform a global search in the space of candidate solutions. This book emphasizes the importance of discovering comprehensible, interesting knowledge, which is potentially useful for intelligent decision making. The text explains both basic concepts and advanced topics

Machine Learning and Knowledge Discovery for Engineering Systems Health Management

Data Mining and Knowledge Discovery with Evolutionary Algorithms

Knowledge Discovery in Multiple Databases

Improving Knowledge Discovery in Science, Engineering, and Business

21st Pacific-Asia Conference, PAKDD 2017, Jeju, South Korea, May 23-26, 2017, Proceedings, Part I

The Data Bonanza

Temporal data mining deals with the harvesting of useful information from temporal data. New initiatives in health care and business organizations have increased the importance of temporal information in data today. From basic data mining concepts to state-of-the-art advances, Temporal Data Mining covers the theory of this subject as well as its application in a variety of fields. It discusses the incorporation of temporality in databases as well as temporal data representation, similarity computation, data classification, clustering, pattern discovery, and prediction. The book also explores the use of temporal data mining in medicine and biomedical informatics, business and industrial applications, web usage mining, and spatiotemporal data mining. Along with various state-of-the-art algorithms, each chapter includes detailed references and short descriptions of relevant algorithms and techniques described in other references. In the appendices, the author explains how data mining fits the overall goal of an organization and how these data can be interpreted for the purpose of characterizing a population. She also provides programs written in the Java language that implement some of the algorithms presented in the first chapter. Check out the author's blog at <http://theophanomitsa.wordpress.com/>

This volume provides an overview of multimedia data mining and knowledge discovery and discusses the variety of hot topics in multimedia data mining research. It describes the objectives and current tendencies in multimedia data mining research and their applications. Each part contains an overview of its chapters and leads the reader with a structured approach through the diverse subjects in the field.

Data Mining Methods for Knowledge Discovery provides an introduction to the data mining methods that are frequently used in the process of knowledge discovery. This book first elaborates on the fundamentals of each of the data mining methods: rough sets, Bayesian analysis, fuzzy sets, genetic algorithms, machine learning, neural networks, and preprocessing techniques. The book then goes on to thoroughly discuss these methods in the setting of the overall process of knowledge discovery. Numerous illustrative examples and experimental findings are also included. Each chapter comes with an extensive bibliography. Data Mining Methods for Knowledge Discovery is intended for senior undergraduate and graduate students, as well as a broad audience of professionals in computer and information sciences, medical informatics, and business information systems.

The importance of having efficient and effective methods for data mining and knowledge discovery (DM&KD), to which the present book is devoted, grows every day and numerous such methods have been developed in recent decades.

There exists a great variety of different settings for the main problem studied by data mining and knowledge discovery, and it seems that a very popular one is formulated in terms of binary attributes. In this setting, states of nature of the application area under consideration are described by Boolean vectors defined on some attributes. That is, by data points defined in the Boolean space of the attributes. It is postulated that there exists a partition of this space into two classes, which should be inferred as patterns on the attributes when only several data points are known, the so-called positive and negative training examples. The main problem in DM&KD is defined as finding rules for recognizing (classifying) new data points of unknown class, i. e., deciding which of them are positive and which are negative. In other words, to infer the binary value of one more attribute, called the goal or class attribute. To solve this problem, some methods have been suggested which construct a Boolean function separating the two given sets of positive and negative training data points.

Database Support for Data Mining Applications

Statistical Data Analytics

Temporal Data Mining

Feature Selection for Knowledge Discovery and Data Mining

Knowledge Discovery in the Social Sciences

An Introduction to Data Mining

Modern medicine generates, almost daily, huge amounts of heterogeneous data. For example, medical data may contain SPECT images, signals like ECG, clinical information like temperature, cholesterol levels, etc., as well as the physician's interpretation. Those who deal with such data understand that there is a widening gap between data collection and data comprehension. Computerized techniques are needed to help humans address this problem. This volume is devoted to the relatively young and growing field of medical data mining and knowledge discovery. As more and more medical procedures employ imaging as a preferred diagnostic tool, there is a need to develop methods for efficient mining in databases of images. Other significant features are security and confidentiality concerns. Moreover, the physician's interpretation of images, signals, or other technical data, is written in unstructured English which is very difficult to mine. This book addresses all these specific features.

The Essentials of Data Science: Knowledge Discovery Using R presents the concepts of data science through a hands-on approach using free and open source software. It systematically drives an accessible journey through data analysis and machine learning to discover and share knowledge from data. Building on over thirty years' experience in teaching and practising data science, the author encourages a programming-by-example approach to ensure students and practitioners attune to the practise of data science while building their data skills. Proven frameworks are provided as reusable templates. Real world case studies then provide insight for the data scientist to swiftly adapt the templates to new tasks and datasets. The book begins by introducing data science. It then reviews R's capabilities for analysing data by writing computer programs. These programs are developed and explained step by step. From analysing and visualising data, the framework moves on to tried and tested machine learning techniques for predictive modelling and knowledge discovery. Literate programming and a consistent style are a focus throughout the book.

This volume presents state-of-the-art tools and techniques for automatically detecting, diagnosing, and predicting the effects of adverse events in an engineered system. It emphasizes the importance of these techniques in managing the intricate interactions within and between engineering systems to maintain a high degree of reliability. Reflecting the interdisciplinary nature of the field, the book explains how the fundamental algorithms and methods of both physics-based and data-driven approaches effectively address systems health management in application areas such as data centers, aircraft, and software systems.

Data mining is the art and science of intelligent data analysis. By building knowledge from information, data mining adds considerable value to the ever increasing stores of electronic data that abound today. In performing data mining many decisions need to be made regarding the choice of methodology, the choice of data, the choice of tools, and the choice of algorithms. Throughout this book the reader is introduced to the basic concepts and some of the more popular algorithms of data mining. With a focus on the hands-on end-to-end process for data mining, Williams guides the reader through various capabilities of the easy to use, free, and open source Rattle Data Mining Software built on the sophisticated R Statistical Software. The focus on doing data mining rather than just reading about data mining is refreshing. The book covers data understanding, data preparation, data refinement, model building, model evaluation, and practical deployment. The reader will learn to rapidly deliver a data mining project using software easily installed for free from the Internet. Coupling Rattle with R delivers a very sophisticated data mining environment with all the power, and more, of the many commercial offerings.

Geographic Data Mining and Knowledge Discovery

17th International Conference, DaWaK 2015, Valencia, Spain, September 1-4, 2015, Proceedings

The Info-Fuzzy Network (IFN) Methodology

Data Analysis, Machine Learning and Knowledge Discovery

The Essentials of Data Science: Knowledge Discovery Using R

Knowledge Discovery with Support Vector Machines

"This book provides a focal point for research and real-world data mining practitioners that advance knowledge discovery from low-quality data; it presents in-depth experiences and methodologies, providing theoretical and empirical guidance to users who have suffered from underlying low-quality data. Contributions also focus on interdisciplinary collaborations among data quality, data processing, data mining, data privacy, and data sharing"--Provided by publisher.

Complete guidance for mastering the tools and techniques of the digital revolution With the digital revolution opening up tremendous opportunities in many fields, there is a growing need for skilled professionals who can develop data-intensive systems and extract information and knowledge from them. This book frames for the first time a new systematic approach for tackling the challenges of data-intensive computing, providing decision makers and technical experts alike with practical tools for dealing with our exploding data collections. Emphasizing data-intensive thinking and interdisciplinary collaboration, The Data Bonanza: Improving Knowledge Discovery in Science, Engineering, and Business examines the essential components of knowledge discovery, surveys many of the current research efforts worldwide, and points to new areas for innovation. Complete with a wealth of examples and DISPEL-based methods demonstrating how to gain more from data in real-world systems, the book: Outlines the concepts and rationale for implementing data-intensive computing in organizations Covers from the ground up problem-solving strategies for data analysis in a data-rich world Introduces techniques for data-intensive engineering using the Data-Intensive Systems Process Engineering Language DISPEL Features in-depth case studies in customer relations, environmental hazards, seismology, and more Showcases successful applications in areas ranging from astronomy and the humanities to transport engineering Includes sample program snippets throughout the text as well as additional materials on a companion website The Data Bonanza is a must-have guide for information strategists, data analysts, and engineers in business, research, and government, and for anyone wishing to be on the cutting edge of data mining, machine learning, databases, distributed systems, or large-scale computing.

Data analysis, machine learning and knowledge discovery are research areas at the intersection of computer science, artificial intelligence, mathematics and statistics. They cover general methods and techniques that can be applied to a vast set of applications such as web and text mining, marketing, medicine, bioinformatics and business intelligence. This volume contains the revised versions of selected papers in the field of data analysis, machine learning and knowledge discovery presented during the 36th annual conference of the German Classification Society (GfKI). The conference was held at the University of Hildesheim (Germany) in August 2012.

This two-volume set, LNAI 10234 and 10235, constitutes the thoroughly refereed proceedings of the 21st Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining, PAKDD 2017, held in Jeju, South Korea, in May 2017. The 129 full papers were carefully reviewed and selected from 458 submissions. They are organized in topical sections named: classification and deep learning; social network and graph mining; privacy-preserving mining and security/risk applications; spatio-temporal and sequential data mining; clustering and anomaly detection; recommender system; feature selection; text and opinion mining; clustering and matrix factorization; dynamic, stream data mining; novel models and algorithms; behavioral data mining; graph clustering and community detection; dimensionality reduction.

Mining the Web

Foundations for Data Mining, Informatics, and Knowledge Discovery, Solutions Manual

Knowledge Discovery from Data Streams

Data Mining and Knowledge Discovery for Geoscientists

Soft Computing for Knowledge Discovery and Data Mining

Statistical Data Mining and Knowledge Discovery

An easy-to-follow introduction to support vector machines This book provides an in-depth, easy-to-follow introduction to support vector machines drawing only from minimal, carefully motivated technical and mathematical background material. It begins with a cohesive discussion of machine learning and goes on to cover: Knowledge discovery environments Describing data mathematically Linear decision surfaces and functions Perceptron learning Maximum margin classifiers Support vector machines Elements of statistical learning theory Multi-class classification Regression with support vector machines Novelty detection Complemented with hands-on exercises, algorithm descriptions, and data sets, Knowledge Discovery with Support Vector Machines is an invaluable textbook for advanced undergraduate and graduate courses. It is also an excellent tutorial on support vector machines for professionals who are pursuing research in machine learning and related areas.

Discovering Knowledge in DataAn Introduction to Data MiningJohn Wiley & SonsData Mining and Knowledge Discovery for GeoscientistsElsevier

Since the beginning of the Internet age and the increased use of ubiquitous computing devices, the large volume and continuous flow of distributed data have imposed new constraints on the design of learning algorithms. Exploring how to extract knowledge structures from evolving and time-changing data, Knowledge Discovery from Data Streams presents a coherent overview of state-of-the-art research in learning from data streams. The book covers the fundamentals that are imperative to understanding data streams and describes important applications, such as TCP/IP traffic, GPS data, sensor networks, and customer click streams. It also addresses several challenges of data mining in the future, when stream mining will be at the core of many applications. These challenges involve designing useful and efficient data mining solutions applicable to real-world problems. In the appendix, the author includes examples of publicly available software and online data sets. This practical, up-to-date book focuses on the new requirements of the next generation of data mining. Although the concepts presented in the text are mainly about data streams, they also are valid for different areas of machine learning and data mining.

This comprehensive textbook on data mining details the unique steps of the knowledge discovery process that prescribes the sequence in which data mining projects should be performed, from problem and data understanding through data preprocessing to deployment of the results. This knowledge discovery approach is what distinguishes Data Mining from other texts in this area. The book provides a suite of exercises and includes links to instructional presentations. Furthermore, it contains appendices of relevant mathematical material.

Privacy-Aware Knowledge Discovery

Knowledge Discovery from Sensor Data

Data Mining Methods for Knowledge Discovery

Big Data Analytics and Knowledge Discovery

Scientific Data Mining and Knowledge Discovery

Data Mining with Rattle and R

As computer power grows and data collection technologies advance, a plethora of data is generated in almost every field where computers are used. The computer generated data should be analyzed by computers; without the aid of computing technologies, it is certain that huge amounts of data collected will not ever be examined, let alone be used to our advantages. Even with today's advanced computer technologies (e. g. , machine learning and data mining systems), discovering knowledge from data can still be fiendishly hard due to the characteristics of the computer generated data. Taking its simplest form, raw data are represented in feature-values. The size of a dataset can be measured in two dimensions, number of features (N) and number of instances (P). Both N and P can be enormously large. This enormity may cause serious problems to many data mining systems. Feature selection is one of the long existing methods that deal with these problems. Its objective is to select a minimal subset of features according to some reasonable criteria so that the original task can be achieved equally well, if not better. By choosing a minimal subset of features, irrelevant and redundant features are removed according to the criterion. When N is reduced, the data space shrinks and in a sense, the data set is now a better representative of the whole data population. If necessary, the reduction of N can also give rise to the reduction of P by eliminating duplicates.

The Definitive Volume on Cutting-Edge Exploratory Analysis of Massive Spatial and Spatiotemporal DatabasesSince the publication of the first edition of Geographic Data Mining and Knowledge Discovery, new techniques for geographic data warehousing (GDW), spatial data mining, and geovisualization (GVis) have been developed. In addition, there has been

Massive data sets pose a great challenge to many cross-disciplinary fields, including statistics. The high dimensionality and different data types and structures have now outstripped the capabilities of traditional statistical, graphical, and data visualization tools. Extracting useful information from such large data sets calls for novel approaches

Data Mining and Knowledge Discovery Handbook organizes all major concepts, theories, methodologies, trends, challenges and applications of data mining (DM) and knowledge discovery in databases (KDD) into a coherent and unified repository. This book first surveys, then provides comprehensive yet concise algorithmic descriptions of methods, including classic methods plus the extensions and novel methods developed recently. This volume concludes with in-depth descriptions of data mining applications in various interdisciplinary industries including finance, marketing, medicine, biology, engineering, telecommunications, software, and security. Data Mining and Knowledge Discovery Handbook is designed for research scientists and graduate-level students in computer science and engineering. This book is also suitable for professionals in fields such as computing applications, information systems management, and strategic research management.

Decomposition Methodology for Knowledge Discovery and Data Mining

Principles and Foundations

Trends and Applications in Knowledge Discovery and Data Mining

A Data Mining Approach

Data Mining and Knowledge Discovery via Logic-Based Methods

Theory and Applications

Data mining from traditional relational databases as well as from non-traditional ones such as semi-structured data, Web data, and scientific databases housing biological, linguistic, and sensor data has recently become a popular way of discovering hidden knowledge. This book on database support for data mining is developed to approaches exploiting the available database technology, declarative data mining, intelligent querying, and associated issues, such as optimization, indexing, query processing, languages, and constraints. Attention is also paid to the solution of data preprocessing problems, such as data cleaning, discretization, and sampling. The 16 reviewed full papers presented were carefully selected from various workshops and conferences to provide complete and competent coverage of the core issues. Some papers were developed within an EC funded project on discovering knowledge with inductive queries.

Knowledge Discovery in Big Data from Astronomy and Earth Observation: Astrogeoinformatics bridges the gap between astronomy and geoscience in the context of applications, techniques and key principles of big data. Machine learning and parallel computing are increasingly becoming cross-disciplinary as the phenomena of Big Data is becoming common place. This book provides insight into the common workflows and data science tools used for big data in astronomy and geoscience. After establishing similarity in data gathering, pre-processing and handling, the data science aspects are illustrated in the context of both fields. Software, hardware and algorithms of big data are addressed. Finally, the book offers insight into the emerging science which combines data and expertise from both fields in studying the effect of cosmos on the earth and its inhabitants. Addresses both astronomy and geosciences in parallel, from a big data perspective Includes introductory information, key principles, applications and the latest techniques Well-supported by computing and information science-oriented chapters to introduce the necessary knowledge in these fields

Mohamed Medhat Gaber "It is not my aim to surprise or shock you - but the simplest way I can summarise is to say that there are now in the world machines that think, that learn and that create.

Moreover, their ability to do these things is going to increase rapidly until - in a visible future - the range of problems they can handle will be coextensive with the range to which the human mind has been applied" by Herbert A. Simon (1916-2001) 1Overview This book suits both graduate students and researchers with a focus on discovering knowledge from scientific data. The use of computational power for data analysis and knowledge discovery in scientific disciplines has found its roots with the evolution of high-performance computing systems. Computational science in physics, chemistry, and biology represents the first step towards automation of data analysis tasks. The rationale behind the development of computational science in different areas was automating mathematical operations performed in those areas. There was no attention paid to the scientific discovery process. Automated Scientific Discovery (ASD) [1-3] represents the second natural step. ASD attempted to automate the process of theory discovery supported by studies in philosophy of science and cognitive sciences. Although early research articles have shown great successes, the area has not evolved due to many reasons. The most important reason was the lack of interaction between scientists and the automating systems.

This book constitutes the refereed proceedings of the 17th International Conference on Data Warehousing and Knowledge Discovery, DaWaK 2015, held in Valencia, Spain, September 2015. The 31 revised full papers presented were carefully reviewed and selected from 90 submissions. The papers are organized in topical sections similarity measure and clustering; data mining; social computing; heterogeneous networks and data; data warehouses; stream processing; applications of big data analysis; and big data.

Knowledge Discovery in Big Data from Astronomy and Earth Observation

Theory, Algorithms, and Applications

Urban Informatics

Knowledge Discovery and Data Mining

Discovering Knowledge with Inductive Queries

Data Mining and Knowledge Discovery Handbook

Many organizations have an urgent need of mining their multiple databases inherently distributed in branches (distributed data). In particular, as the Web is rapidly becoming an information flood, individuals and organizations can take into account low-cost information and knowledge on the Internet when making decisions. How to efficiently identify quality knowledge from different data sources has become a significant challenge. This challenge has attracted a great many researchers including the authors who have developed a local pattern analysis, a new strategy for discovering some kinds of potentially useful patterns that cannot be mined in traditional multi-database mining techniques. Local pattern analysis delivers high-performance pattern discovery from multiple databases. There has been considerable progress made on multi-database mining in such areas as hierarchical meta-learning, collective mining, database classification, and peculiarity discovery. While these techniques continue to be future topics of interest concerning multi-database mining, this book focuses on these interesting issues under the framework of local pattern analysis. The book is intended for researchers and students in data mining, distributed data analysis, machine learning, and anyone else who is interested in multi-database mining. It is also appropriate for use as a text supplement for broader courses that might also involve knowledge discovery in databases and data mining.

Data Mining is the science and technology of exploring large and complex bodies of data in order to discover useful patterns. It is extremely important because it enables modeling and knowledge extraction from abundant data availability.

This book introduces soft computing methods extending the envelope of problems that data mining can solve efficiently. It presents practical soft-computing approaches in data mining and includes various real-world case studies with detailed results.

Knowledge Discovery in the Social Sciences helps readers find valid, meaningful, and useful information. It is written for researchers and data analysts as well as students who have no prior experience in statistics or computer science.

Suitable for a variety of classes—including upper-division courses for undergraduates, introductory courses for graduate students, and courses in data management and advanced statistical methods—the book guides readers in the application of data mining techniques and illustrates the significance of newly discovered knowledge. Readers will learn to:

- appreciate the role of data mining in scientific research
- develop an understanding of fundamental concepts of data mining and knowledge discovery
- use software to carry out data mining tasks
- select and assess appropriate models to ensure findings are valid and meaningful
- develop basic skills in data preparation, data mining, model selection, and validation
- apply concepts with end-of-chapter exercises and review summaries

Knowledge Discovery and Data Mining: Challenges and Realities

Information Visualization in Data Mining and Knowledge Discovery

Novel Applications and New Techniques

Astrogeoinformatics