Free read 2013 ieee paper on data stream mining [PDF]

this book is a significant contribution to the subject of mining time changing data streams and addresses the design of learning algorithms for this purpose it introduces new contributions on several different aspects of the problem identifying research opportunities and increasing the scope for applications it also includes an in depth study of stream mining and a theoretical analysis of proposed methods and algorithms the first section is concerned with the use of an adaptive sliding window algorithm adwin since this has rigorous performance guarantees using it in place of counters or accumulators it offers the possibility of extending such guarantees to learning and mining algorithms not initially designed for drifting data testing with several methods including naive bayes clustering decision trees and ensemble methods is discussed as well the second part of the book describes a formal study of connected acyclic graphs or trees from the point of view of closure based mining presenting efficient algorithms for subtree testing and for mining ordered and unordered frequent closed trees lastly a general methodology to identify closed patterns in a data stream is outlined this is applied to develop an incremental method a sliding window based method and a method that mines closed trees adaptively from data streams these are used to introduce classification methods for tree data streams 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 in recent years data streams became a ubiquitous source of information and thus stream mining emerged as a new field in database research due to the inherently dynamic nature of data streams stream mining algorithms benefit from being adaptive to changes in the properties of a data stream in addition when stream mining is done in a dynamic environment like a data stream management system or a sensor network stream mining algorithms also profit from being adaptive to the changing conditions in this environment this work investigates two kinds of adaptivity in data stream mining first a model for quality driven resource adaptive stream mining is developed the model is applied to stream mining algorithms so they efficiently utilize available resources to achieve mining results of the highest quality possible every stream mining algorithm is unique in its parameters quality measures and resource consumption patterns we generalize these characteristics and develop a model that captures the interactions and correlations between variables involved in the stream mining process we then express resource adaptive stream mining as a multiobjective optimization problem and use its solution to tune the input parameters of stream mining algorithms which results in high quality mining and optimal resource utilization the second topic investigated in this work is feature adaptive stream mining which is concerned with adjusting the focus of the mining process to interesting features detected in the data stream this research is motivated by the need to efficiently detect environmental phenomena from sensor data streams we propose methods to detect and predict heterogeneous outlier regions which represent areas of environmental phenomena of different intensities with the help of predictions about the location and size of outlier regions the sampling rate of individual sensors is adapted such that sensors in the vicinity of environmental phenomena obtain new measurements more frequently than other sensors network to allow for a precise and timely region tracking the research in this work enhances the state of the art in data stream mining as it makes stream mining algorithms more flexible to adapt to changes in the data stream and the mining environment this volume focuses on the theory and practice of data stream management and the novel challenges this emerging domain poses for data management
algorithms systems and applications the collection of chapters contributed by authorities in the field offers a comprehensive introduction to both the algorithmic theoretical foundations of data streams as well as the streaming systems and applications built in different domains a short introductory chapter provides a brief summary of some basic data streaming concepts and models and discusses the key elements of a generic stream query processing architecture subsequently part i focuses on basic streaming algorithms for some key analytics functions e.g. quantiles norms join aggregates heavy hitters over streaming data part ii then examines important techniques for basic stream mining tasks e.g. clustering classification frequent itemsets part iii discusses a number of advanced topics on stream processing algorithms and part iv focuses on system and language aspects of data stream processing with surveys of influential system prototypes and language designs part v then presents some representative applications of streaming techniques in different domains e.g. network management financial analytics finally the volume concludes with an overview of current data streaming products and new application domains e.g. cloud computing big data analytics and complex event processing and a discussion of future directions in this exciting field the book provides a comprehensive overview of core concepts and technological foundations as well as various systems and applications and is of particular interest to students and researchers in the area of data stream management this book constitutes the proceedings of the third international conference on data stream and mining and processing dsmp 2020 held in lviv ukraine in august 2020 the 36 full papers presented in this volume were carefully reviewed and selected from 134 submissions the papers are organized in topical sections of hybrid systems of computational intelligence machine vision and pattern recognition dynamic data mining data stream mining big data data science using intelligent approaches the conference was held virtually due to the covid 19 pandemic 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 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 the types of risks that may impact data stream mining been identified and analyzed are there any easy to implement alternatives to data stream mining sometimes other solutions are available that do not require the cost implications of a full blown project is there a critical path to deliver data stream mining results what prevents me from making the changes i know will make me a more effective data stream mining leader who is the data stream mining process owner this valuable data stream mining self assessment will make you the credible data stream mining domain master by revealing just what you need to know to be fluent and ready for any data stream mining challenge how do i reduce the effort in the data stream mining work to
be done to get problems solved how can i ensure that plans of action include every data stream mining task and that every data stream mining outcome is in place how will i save time investigating strategic and tactical options and ensuring data stream mining costs are low how can i deliver tailored data stream mining advice instantly with structured going forward plans there's no better guide through these mind expanding questions than acclaimed best selling author gerard blokdyk blokdyk ensures all data stream mining essentials are covered from every angle the data stream mining self assessment shows succinctly and clearly that what needs to be clarified to organize the required activities and processes so that data stream mining outcomes are achieved contains extensive criteria grounded in past and current successful projects and activities by experienced data stream mining practitioners their mastery combined with the easy elegance of the self assessment provides its superior value to you in knowing how to ensure the outcome of any efforts in data stream mining are maximized with professional results your purchase includes access details to the data stream mining self assessment dashboard download which gives you your dynamically prioritized projects ready tool and shows you exactly what to do next your exclusive instant access details can be found in your book what would be the goal or target for a data stream mining s improvement team are there recognized data stream mining problems how will the data stream mining team and the organization measure complete success of data stream mining is the impact that data stream mining has shown are accountability and ownership for data stream mining clearly defined this amazing data stream mining self assessment will make you the credible data stream mining domain veteran by revealing just what you need to know to be fluent and ready for any data stream mining challenge how do i reduce the effort in the data stream mining work to be done to get problems solved how can i ensure that plans of action include every data stream mining task and that every data stream mining outcome is in place how will i save time investigating strategic and tactical options and ensuring data stream mining costs are low how can i deliver tailored data stream mining advice instantly with structured going forward plans there's no better guide through these mind expanding questions than acclaimed best selling author gerard blokdyk blokdyk ensures all data stream mining essentials are covered from every angle the data stream mining self assessment shows succinctly and clearly that what needs to be clarified to organize the required activities and processes so that data stream mining outcomes are achieved contains extensive criteria grounded in past and current successful projects and activities by experienced data stream mining practitioners their mastery combined with the easy elegance of the self assessment provides its superior value to you in knowing how to ensure the outcome of any efforts in data stream mining are maximized with professional results your purchase includes access details to the data stream mining self assessment dashboard download which gives you your dynamically prioritized projects ready tool and shows you exactly what to do next your exclusive instant access details can be found in your book this compendium is a completely revised version of an earlier book data mining in time series databases by the same editors it provides a unique collection of new articles written by leading experts that account for the latest developments in the field of time series and data stream mining the emerging topics covered by the book include weightless neural modeling for mining data streams using ensemble classifiers for imbalanced and evolving data streams document stream mining with active learning and many more in particular it addresses the domain of streaming data which has recently become one of the emerging topics in data science big data and related areas existing titles do not provide sufficient information on this topic contents streaming data mining with massive online analytics moa albert bifet jesse read geoff holmes and bernhard pfahringer weightless neural modeling for mining data streams douglas o cardoso joão gama and felipe frança ensemble classifiers for imbalanced and evolving data streams dariusz brzeziński and jerzy stefanowski consensus learning for sequence data andreas nienkötter and xiaoyi jiang clustering based classification of document streams with active learning mark last maxim stolar and menahem friedman supporting the mining of big data by means of domain knowledge during the pre mining phases rémon cornelisse and sunil choenni data
analytics industrial perspective solutions for streaming data mohsin munir sebastian baumbach ying gu andreas dengel and sheraz ahmed
readership researchers academics professionals and graduate students in artificial intelligence machine learning databases and information
science keywords time series data streams big data internet of things concept drift sequence mining episode mining incremental learning
active learning review 0 this dissertation techniques in data stream mining by suk man ivy tong 湯淑敏 was obtained from the university of hong
kong pokfulam hong kong and is being sold pursuant to creative commons attribution 3.0 hong kong license the content of this dissertation
has not been altered in any way we have altered the formatting in order to facilitate the ease of printing and reading of the dissertation
all rights not granted by the above license are retained by the author abstract abstract of thesis entitled techniques in data stream
mining submitted by tong suk man ivy for the degree of master of philosophy at the university of hong kong in november 2005 many
organizations have been confronted by a data explosion in the last decade and face the problem of managing very large databases that grow
at a rate of several million records per day to address this problem database and data mining communities have recently focused on stream
processing where data arrives in the form of continuous data streams efficient stream mining is challenging yet critical however it is not
feasible to perform traditional data mining algorithms on streaming data is infeasible they have a number of limitations 1 most of the
classical mining algorithms take multiple passes over the entire database but the speed of arrival and the volume of the data streams makes
it impossible to store them 2 timely response is important in stream applications disk based algorithms are inappropriate 3 since only a
small representation of the whole dataset is kept approximate algorithms with high accuracy are needed this studyexplores some techniques
in data stream mining in particular it focuses on data from multiple sensor streams where each stream represents a sequence of states of a
monitored attribute reported by a sensor against time in finance a stream may be a stock for example the first technique proposed in this
study is a modification of vitter 's reservoir sampling algorithm which can generate a fixed size uniform sample set from an input stream
without a priori knowledge of the size of the stream applying reservoir sampling on each stream individually would give a sample of time
uncorrelated points from different sensor streams that is the sensor states sampled for different streams do not co exist within any time
span the sample obtained is therefore useless for answering queries related to associations of the streams instead of sampling streams
individually a sample of snapshots taken of the stream at different time instants is generated this ensures that if the state of a stream
in a certain time span is sampled the states of other streams in the time span must be in the sample the second technique is used in mining
frequent patterns from a large sensor network data representation of sensor streams affects the efficiency of online mining based on the
estimation mechanism of the lossy counting lc algorithm a window based algorithm ilb which makes use of interval list representation is
proposed experiments on synthetic datasets were conducted to show the efficiency of our ilb algorithms experimental results showed that if
the sensor network is massive the ilb algorithms outperform lc by a significant margin doi 10.5353/th/b3473737 subjects database management
data mining this book presents a unique approach to stream data mining unlike the vast majority of previous approaches which are largely
based on heuristics it highlights methods and algorithms that are mathematically justified first it describes how to adapt static decision
trees to accommodate data streams in this regard new splitting criteria are developed to guarantee that they are asymptotically equivalent
to the classical batch tree moreover new decision trees are designed leading to the original concept of hybrid trees in turn nonparametric
techniques based on parzen kernels and orthogonal series are employed to address concept drift in the problem of non stationary regressions
and classification in a time varying environment lastly an extremely challenging problem that involves designing ensembles and
automatically choosing their sizes is described and solved given its scope the book is intended for a professional audience of researchers
and practitioners who deal with stream data e g in telecommunication banking and sensor networks processing data streams has raised new
research challenges over the last few years this book provides the reader with a comprehensive overview of stream data processing including famous prototype implementations like the nile system and the tinyos operating system applications in security the natural sciences and education are presented the huge bibliography offers an excellent starting point for further reading and future research knowledge discovery and data mining from time changing data streams and concept drift handling on data streams have become important topics in the machine learning recently machine learning offers promise of a solution but the field mainly focuses on achieving high accuracy when data supply is limited the challenges that are faced by information processing and classification in particular are related to the need to cope with huge volume of data to process data streams online and in real time and to handle concept drift when tackling with data stream incremental classification algorithms are required an ensemble of classifiers has several advantages over single classifier methods so we have designed and implemented a new ensemble classifier which is adaptive and efficient for data streams classification adaptive sliding window and adaptive size hoeffding tree techniques are used in this algorithm this technique should helpful to online processing of data streams and should be especially useful to network monitoring systems and financial industries or anyone else who may be handling data streams data stream mining a complete guide this springerbrief presents the fundamental concepts of a specialized class of data stream spatio temporal data streams and demonstrates their distributed processing using big data frameworks and platforms it explores a consistent framework which facilitates a thorough understanding of all different facets of the technology from basic definitions to state of the art techniques key topics include spatio temporal continuous queries distributed stream processing sql like language embedding and trajectory stream clustering over the course of the book the reader will become familiar with spatio temporal data streams management and data flow processing which enables the analysis of huge volumes of location aware continuous data streams applications range from mobile object tracking and real time intelligent transportation systems to traffic monitoring and complex event processing spatio temporal data streams is a valuable resource for researchers studying spatio temporal data streams and big data applications as well as data engineers and data scientists solving data management and analytics problems associated with this class of data data mining ist ein inzwischen etabliertes erfolgereiches werkzeug zur extraktion von neuem bislang unbekanntem wissen aus daten in mittlerweile fast allen gr en Unternehmen wird es genutzt um mehrwerte f r kunden zu generieren den erfolg von marketingkampagnen zu erh hen betrugsverdacht aufzudecken oder beispielsweise durch segmentierung unterschiedliche kundengruppen zu identifizieren ein grundproblem der intelligenten datenanalyse besteht darin dass daten oftmals in rasanter geschwindigkeit neu entstehen eink ufe im supermarkt telefonverbindungen oder der ffentliche verkehr erzeugen tglich eine neue flut an daten in denen potentiell wertvolles wissen steckt die versteckten Zusammenhangen nge und muster k nnen sich im Zeitverlauf mehr oder weniger stark ver ndern und zeitbedingt neue datenmodelemente finden kann oder nicht mehr relevant sind data mining modelle daher schnell veraltet betrugsversuche k nnen dann unter umst nden nicht mehr erkannt und abgewendet werden oder produktionskompetenzen basieren auf veralteten Datenbanken oder aber auch auf veralteten Methoden die neuen Datenmodelle und Methoden sind daher notwendig um den Wettbewerbsvorteil zu behalten geoinformatics and data mining focuses on the development and implementation of computational algorithms to solve these problems this unique volume contains a collection of chapters on state of the art data mining techniques applied to geoinformatics problems of high complexity and important societal value data mining for geoinformatics addresses current concerns and developments relating to
spatio temporal data mining issues in remotely sensed data problems in meteorological data such as tornado formation estimation of radiation from the fukushima nuclear power plant simulations of traffic data using openstreetmap real time traffic applications of data stream mining visual analytics of traffic and weather data and the exploratory visualization of collective mobile objects such as the flocking behavior of wild chickens this book is designed for researchers and advanced level students focused on computer science earth science and geography as a reference or secondary text book practitioners working in the areas of data mining and geoscience will also find this book to be a valuable reference 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 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 scienti c data the use of computational power for data analysis and knowledge discovery in scienti c disciplines has found its roots with the re lution of high performance computing systems computational science in physics chemistry and biology represents the rst step towards automation of data analysis tasks the rational behind the developmentof computationalscience in different eas was automating mathematical operations performed in those areas there was no attention paid to the scienti c discovery process automated scienti c disc ery 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 the real power for security applications will come from the synergy of academic and commercial research focusing on the specific issue of security special constraints apply to this domain which are not always taken into consideration by academic research but are critical for successful security applications large volumes techniques must be able to handle huge amounts of data and perform on line computation scalability algorithms must have processing times that scale well with ever growing volumes automation the analysis process must be automated so that information extraction can run on its own ease of use everyday citizens should be able to extract and assess the necessary information and robustness systems must be able to cope with data of poor quality missing or erroneous data the nato advanced study institute asi on mining massive data sets for security held in italy september 2007 brought together around ninety participants to discuss these issues this publication includes the most important contributions but can of course not entirely reflect the lively interactions which allowed the participants to exchange their views and share their experience the bridge between academic methods and industrial constraints is systematically discussed throughout this
volume will thus serve as a reference book for anyone interested in understanding the techniques for handling very large data sets and how to apply them in conjunction for solving security issues. Computational intelligence, data mining, data stream mining, machine vision, pattern recognition, internet usage has become a normal and essential aspect of everyday life due to the immense amount of information available on the web. It has become obligatory to find ways to sift through and categorize the overload of data while removing redundant material. Collaborative filtering using data mining and analysis evaluates the latest patterns and trending topics in the utilization of data mining tools and filtering practices featuring emergent research and optimization techniques in the areas of opinion mining, text mining, and sentiment analysis. As well as their various applications, this book is an essential reference source for researchers and engineers interested in collaborative filtering. Data stream mining is becoming very important in many application areas such as the stock market, network traffic, web logs, and ATM transactions. A data stream consists of an ordered sequence of instances and because there are usually a large number of instances along with limited computing and storage capabilities, algorithms that read the data only once are preferred. There has been some research that focuses on finding when a concept has changed given some knowledge about the previous instances in the data stream. But little on determining the characteristics of that change. In this thesis, we concentrate on finding the characteristics of the changes that occur using frequent itemset mining techniques. We propose two approaches both combining heuristic and statistical approaches to analyze changes that have occurred within a stream at itemset level. Hoeffding bound and Bernstein bound inequality are two statistical methods used in our research in order to handle the infinite length and limited labelled data streams. Our approaches incorporate a sliding window algorithm and a couple of effective data structure such as CP tree, CAN tree, and SPO tree to identify three types of change: extension, reduction, and support fluctuation. To experiment and evaluate our algorithms, we customized the IBM Quest Synthetic Data Generator to create types of dataset containing those three types of change. Our algorithms have been fully tested on both synthetic and real-world datasets. Data mining and machine learning applications are elaborated in detail on the current needs of data mining and machine learning and promotes mutual understanding among research in different disciplines. The latest currency of today's world is the new gold. In this new form of gold, the most beautiful jewels are data analytics and machine learning. Data mining and machine learning are considered interdisciplinary fields. Data mining is a subset of data analytics and machine learning. Involves the use of algorithms that automatically improve through experience based on data. Massive datasets can be classified and clustered to obtain accurate results. The most common technologies used include classification and clustering methods. Accuracy and error rates are calculated for regression and classification. Clustering to find actual results through algorithms like support vector machines and neural networks with forward and backward propagation applications include fraud detection, image processing, medical diagnosis, weather prediction, e-commerce, and so forth. The book features a review of the state of the art in data mining and machine learning, a review and description of the learning methods in human-computer interaction, implementation strategies, and future research directions. Used to meet the design and application requirements of several modern and real-time applications, for a long time the scope and implementation of a majority of data mining and machine learning strategies are discussed. Real-time problems, audience industry, and academic researchers, scientists, and engineers in information technology, data science, and machine and deep learning as well as artificial intelligence more broadly.
Adaptive Stream Mining 2010 this book is a significant contribution to the subject of mining time changing data streams and addresses the design of learning algorithms for this purpose it introduces new contributions on several different aspects of the problem identifying research opportunities and increasing the scope for applications it also includes an in depth study of stream mining and a theoretical analysis of proposed methods and algorithms the first section is concerned with the use of an adaptive sliding window algorithm adwin since this has rigorous performance guarantees using it in place of counters or accumulators it offers the possibility of extending such guarantees to learning and mining algorithms not initially designed for drifting data testing with several methods including naïve bayes clustering decision trees and ensemble methods is discussed as well the second part of the book describes a formal study of connected acyclic graphs or trees from the point of view of closure based mining presenting efficient algorithms for subtree testing and for mining ordered and unordered frequent closed trees lastly a general methodology to identify closed patterns in a data stream is outlined this is applied to develop an incremental method a sliding window based method and a method that mines closed trees adaptively from data streams these are used to introduce classification methods for tree data streams

Ein Framework für dynamische, hochgradig parallel arbeitende Data Stream Mining Strukturen 2010 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

Data Streams 2007-04-03 in recent years data streams became a ubiquitous source of information and thus stream mining emerged as a new field in database research due to the inherently dynamic nature of data streams stream mining algorithms benefit from being adaptive to changes in the properties of a data stream in addition when stream mining is done in a dynamic environment like a data stream management system a sensor network stream mining algorithms also profit from being adaptive to the changing conditions in this environment this work investigates two kinds of adaptivity in data stream mining first a model for quality driven resource adaptive stream mining is developed the model is applied to stream mining algorithms so they efficiently utilize available resources to achieve mining results of the highest quality possible every stream mining algorithm is unique in its parameters quality measures and resource consumption patterns we generalize these characteristics and develop a model that captures the interactions and correlations between variables involved in the stream mining process we then express resource adaptive stream mining as a multiobjective optimization problem and use its solution to tune the input parameters of stream mining algorithms which results in high quality mining and optimal resource utilization the second topic investigated in this work is feature adaptive stream mining which is concerned with adjusting the focus of the mining process to interesting features detected in the data stream this research is motivated by the need to efficiently detect environmental phenomena from sensor data streams we propose methods to detect and predict heterogeneous outlier regions which represent areas of environmental phenomena of different intensities with the help of predictions about the location and size of outlier regions the sampling rate of individual sensors is adapted such that sensors in the vicinity of environmental phenomena obtain new measurements more frequently than other sensors in the network to allow for a precise and timely region tracking the research in this work enhances the state of the art in data stream mining as it makes stream mining algorithms more flexible to adapt to changes in the data stream and the mining environment

Adaptivity in Data Stream Mining 2009 this volume focuses on the theory and practice of data stream management and the novel challenges this emerging domain poses for data management algorithms systems and applications the collection of chapters contributed by authorities in
the field offers a comprehensive introduction to both the algorithmic theoretical foundations of data streams as well as the streaming systems and applications built in different domains. A short introductory chapter provides a brief summary of some basic data streaming concepts and models and discusses the key elements of a generic stream query processing architecture. Subsequently, Part I focuses on basic streaming algorithms for some key analytics functions, e.g., quantiles, norms, join, aggregates, heavy hitters over streaming data. Part II then examines important techniques for basic stream mining tasks, e.g., clustering, classification, frequent itemsets. Part III discusses a number of advanced topics on stream processing algorithms, and Part IV focuses on system and language aspects of data stream processing with surveys of influential system prototypes and language designs. Part V then presents some representative applications of streaming techniques in different domains, e.g., network management, financial analytics. Finally, the volume concludes with an overview of current data stream mining products and new application domains, e.g., cloud computing. Big data analytics and complex event processing and a discussion of future directions in this exciting field. The book provides a comprehensive overview of core concepts and technological foundations as well as various systems and applications and is of particular interest to students, lecturers, and researchers in the area of data stream management.

Konzept für ein flexibles Data-Stream-Mining-Framework auf Android 2016

This book constitutes the proceedings of the third international conference on data stream and mining and processing DSMP 2020 held in Lviv, Ukraine in August 2020. The 36 full papers presented in this volume were carefully reviewed and selected from 134 submissions. The papers are organized in topical sections of hybrid systems of computational intelligence, machine vision, and pattern recognition, dynamic data mining, data stream mining, big data, data science, using intelligent approaches. The conference was held virtually due to the COVID-19 pandemic.

Data Stream Management 2016-07-11

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. The book 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 and MoA-based lab sessions. 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.

Data Stream Mining & Processing 2020-11-04

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 challenges.

Machine Learning for Data Streams 2023-05-09

Have the types of risks that may impact data stream mining been identified and analyzed? Are there any easy-to-implement alternatives to data stream mining? Sometimes, other solutions are available that do not require the cost implications of a full-blown project. Is there a critical path to deliver data stream mining results? What prevents me from making the changes I know will make me a more effective data stream mining leader who is the data stream mining process owner? This valuable data
stream mining self assessment will make you the credible data stream mining domain master by revealing just what you need to know to be fluent and ready for any data stream mining challenge how do i reduce the effort in the data stream mining work to be done to get problems solved how can i ensure that plans of action include every data stream mining task and that every data stream mining outcome is in place how will i save time investigating strategic and tactical options and ensuring data stream mining costs are low how can i deliver tailored data stream mining advice instantly with structured going forward plans there s no better guide through these mind expanding questions than acclaimed best selling author gerard blokdyk blokdyk ensures all data stream mining essentials are covered from every angle the data stream mining self assessment shows succinctly and clearly that what needs to be clarified to organize the required activities and processes so that data stream mining outcomes are achieved contains extensive criteria grounded in past and current successful projects and activities by experienced data stream mining practitioners their mastery combined with the easy elegance of the self assessment provides its superior value to you in knowing how to ensure the outcome of any efforts in data stream mining are maximized with professional results your purchase includes access details to the data stream mining self assessment dashboard download which gives you your dynamically prioritized projects ready tool and shows you exactly what to do next your exclusive instant access details can be found in your book Knowledge Discovery from Data Streams 2010-05-25 what would be the goal or target for a data stream mining s improvement team are there recognized data stream mining problems how will the data stream mining team and the organization measure complete success of data stream mining is the impact that data stream mining has shown are accountability and ownership for data stream mining clearly defined this amazing data stream mining self assessment will make you the credible data stream mining domain veteran by revealing just what you need to know to be fluent and ready for any data stream mining challenge how do i reduce the effort in the data stream mining work to be done to get problems solved how can i ensure that plans of action include every data stream mining task and that every data stream mining outcome is in place how will i save time investigating strategic and tactical options and ensuring data stream mining costs are low how can i deliver tailored data stream mining advice instantly with structured going forward plans there s no better guide through these mind expanding questions than acclaimed best selling author gerard blokdyk blokdyk ensures all data stream mining essentials are covered from every angle the data stream mining self assessment shows succinctly and clearly that what needs to be clarified to organize the required activities and processes so that data stream mining outcomes are achieved contains extensive criteria grounded in past and current successful projects and activities by experienced data stream mining practitioners their mastery combined with the easy elegance of the self assessment provides its superior value to you in knowing how to ensure the outcome of any efforts in data stream mining are maximized with professional results your purchase includes access details to the data stream mining self assessment dashboard download which gives you your dynamically prioritized projects ready tool and shows you exactly what to do next your exclusive instant access details can be found in your book Data Stream Mining a Complete Guide 2018-04-16 this compendium is a completely revised version of an earlier book data mining in time series databases by the same editors it provides a unique collection of new articles written by leading experts that account for the latest developments in the field of time series and data stream mining the emerging topics covered by the book include weightless neural modeling for mining data streams using ensemble classifiers for imbalanced and evolving data streams document stream mining with active learning and many more in particular it addresses the domain of streaming data which has recently become one of the emerging topics in data science big data and related areas existing titles do not provide sufficient information on this topic contents streaming data mining with massive online analytics moa albert bifet jesse read geoff holmes and bernhard pfahringer weightless neural modeling for mining data streams douglas o cardoso joão gama and felipe frança ensemble classifiers for imbalanced and evolving data streams dariusz brzezinski and jerzy
Data Stream Mining 2018-05-07 this dissertation techniques in data stream mining by suk man ivy tong was obtained from the university of hong kong pokfulam hong kong and is being sold pursuant to creative commons attribution 3.0 hong kong license the content of this dissertation has not been altered in any way we have altered the formatting in order to facilitate the ease of printing and reading of the dissertation all rights not granted by the above license are retained by the author abstract abstract of thesis entitled techniques in data stream mining submitted by tong suk man ivy for the degree of master of philosophy at the university of hong kong in november 2005 many organizations have been confronted by a data explosion in the last decade and face the problem of managing very large databases that grow at a rate of several million records per day to address this problem database and data mining communities have recently focused on stream processing where data arrives in the form of continuous data streams efficient stream mining is challenging yet critical however it is not feasible to perform traditional data mining algorithms on streaming data is infeasible they have a number of limitations 1 most of the classical mining algorithms take multiple passes over the entire database but the speed of arrival and the volume of the data streams makes it impossible to store them 2 timely response is important in stream applications disk based algorithms are inappropriate 3 since only a small representation of the whole dataset is kept approximate algorithms with high accuracy are needed this study explores some techniques in data stream mining in particular it focuses on data from multiple sensor streams where each stream represents a sequence of states of a monitored attribute reported by a sensor against time in finance a stream may be a stock for example the first technique proposed in this study is a modification of vitter’s reservoir sampling algorithm which can generate a fixed size uniform sample set from an input stream without a priori knowledge of the size of the stream applying reservoir sampling on each stream individually would give a sample of time uncorrelated points from different sensor streams that is the sensor states sampled for different streams do not coexist within any time span the sample obtained is therefore useless for answering queries related to associations of the streams instead of sampling streams individually a sample of snapshots taken of the streams at different time instants is generated this ensures that if the state of a stream in a certain time span is sampled the states of other streams in the time span must be in the sample the second technique is used in mining frequent patterns from a large sensor network data representation of sensor streams affects the efficiency of online mining based on the estimation mechanism of the lossy counting lc algorithm a window based algorithm ilb which makes use of interval list representation is proposed experiments on synthetic datasets were conducted to show the efficiency of our ilb algorithms experimental results showed that if the sensor network is massive the ilb algorithms outperform lc by a significant margin doi:10.5353/th/b3473737 subjects database management data mining Data Mining In Time Series And Streaming Databases 2018-01-11 this book presents a unique approach to stream data mining unlike the vast majority of previous approaches which are largely based on heuristics it highlights methods and algorithms that are mathematically justified first it describes how to adapt static decision trees to accommodate data streams in this regard new splitting criteria are developed to guarantee that they are asymptotically equivalent to the classical batch tree moreover new decision trees are designed leading
to the original concept of hybrid trees in turn nonparametric techniques based on parzen kernels and orthogonal series are employed to address concept drift in the problem of non stationary regressions and classification in a time varying environment largely an extremely challenging problem that involves designing ensembles and automatically choosing their sizes is described and solved given its scope the book is intended for a professional audience of researchers and practitioners who deal with stream data e.g. in telecommunication banking and sensor networks.

Techniques in Data Stream Mining 2017-01-26 processing data streams has raised new research challenges over the last few years this book provides the reader with a comprehensive overview of stream data processing including famous prototype implementations like the nile system and the tinyos operating system applications in security the natural sciences and education are presented the huge bibliography offers an excellent starting point for further reading and future research.

Stream Data Mining: Algorithms and Their Probabilistic Properties 2019-03-16 knowledge discovery and data mining from time changing data streams and concept drift handling on data streams have become important topics in the machine learning recently machine learning offers promise of a solution but the field mainly focuses on achieving high accuracy when data supply is limited the challenges that are faced by information processing and classification in particular are related to the need to cope with huge volume of data to process data streams online and in real time and to handle concept drift when tackling with data stream incremental classification algorithms are required an ensemble of classifiers has several advantages over single classifier methods so we have designed and implemented a new ensemble classifier which is adaptive and efficient for data streams classification adaptive sliding window and adaptive size hoeffding tree techniques are used in this algorithm this technique should help in online processing of data streams and should be especially useful to network monitoring systems and financial industries or anyone else who may be handling data streams.

Learning from Data Streams 2007-09-20 data stream mining a complete guide.

Data Streams Mining 2010-12 this springerbrief presents the fundamental concepts of a specialized class of data stream spatio temporal data streams and demonstrates their distributed processing using big data frameworks and platforms it explores a consistent framework which facilitates a thorough understanding of all different facets of the technology from basic definitions to state of the art techniques key topics include spatio temporal continuous queries distributed stream processing sql like language embedding and trajectory stream clustering over the course of the book the reader will become familiar with spatio temporal data streams management and data flow processing which enables the analysis of huge volumes of location aware continuous data streams applications range from mobile object tracking and real time intelligent transportation systems to traffic monitoring and complex event processing spatio temporal data streams is a valuable resource for researchers studying spatio temporal data streams and big data analytics as well as data engineers and data scientists solving data management and analytics problems associated with this class of data.

Data Stream Mining A Complete Guide 2018 data mining is ein inzwischen etabliertes erfolgreiches werkzeug zur extraktion von neuem bislang unbekanntem wissen aus daten in mittlerweile fast allen größen unternehmen wird es genutzt um mehrwerte für kunden zu generieren den erfolg von marketingkampagnen zu erhöhen betrugsverdacht aufzudecken oder beispielsweise durch segmentierung unterschiedliche kundengruppen zu identifizieren ein grundproblem der intelligenten datenanalyse besteht darin dass daten oftmals in rasanter geschwindigkeit neu entstehen einkauf im supermarkt telefonverbindungen oder der öffentliche verkehr erzeugen eine neue flut an daten in denen potentiell wertvolles wissen steckt die versteckten Zusammenhänge und muster k nnen sich im zeitverlauf mehr oder weniger stark verndern datenmodellierung findet in der regel aber noch immer einmalig bzw sporadisch auf dem snapshot einer datenbank statt einmal erkannte muster
oder zusammenhängen werden auch dann noch angenommen wenn diese längst nicht mehr bestehen gerade in dynamischen Umgebungen wie zum Beispiel einem Internet-Shop sind Data-Mining-Modelle daher schnell veraltet. Betrugsversuche können dann unter Umständen nicht mehr erkannt werden oder Absatzpotentiale nicht mehr genutzt werden, oder Produktempfehlungen basieren auf veralteten Warenkörben. Um dauerhaft Wettbewerbsvorteile erzielen zu können, muss das Wissen ber Daten aber möglichst aktuell und von ausgezeichneter Qualität sein. Der Inhalt dieses Buches skizziert Methoden und Vorgehensweisen von Data-Mining in Echtzeit.

**Techniques in Data Stream Mining** 2005. The rate at which geospatial data is being generated exceeds our computational capabilities to extract patterns for the understanding of a dynamically changing world. Geoinformatics and data mining focuses on the development and implementation of computational algorithms to solve these problems. This unique volume contains a collection of chapters on state of the art data mining techniques applied to geoinformatic problems of high complexity and important societal value. Data mining for geoinformatics addresses current concerns and developments relating to spatio-temporal data mining issues in remotely sensed data problems in meteorological data such as tornado formation estimation of radiation from the Fukushima nuclear power plant simulations of traffic data using OpenStreetMap, real-time traffic applications of data stream mining, visual analytics of traffic and weather data, and the exploratory visualization of collective mobile objects such as the flocking behavior of wild chickens. This book is designed for researchers and advanced level students focused on computer science, earth science, and geography as a reference or secondary text book. Practitioners working in the areas of data mining and geoscience will also find this book to be a valuable reference.

**Spatio-Temporal Data Streams** 2016. 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 this book. Potential applications of paramount importance offered by PDM in a variety of domains including security business and telemedicine are discussed.

**Data Stream Mining** 2007. Mohamed Medhat Gaber.

It is 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). An overview of 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 re-emergence of high-performance computing. Systems computational science in physics, chemistry, and biology. Represents the next step towards automation of data analysis tasks the rational 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, as depicted by AS and 13, represents the second natural step as it 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

Statistical and Probabilistic Methods for Data Stream Mining 2012 the real power for security applications will come from the synergy of academic and commercial research focusing on the specific issue of security special constraints apply to this domain which are not always taken into consideration by academic research but are critical for successful security applications large volumes techniques must be able to handle huge amounts of data and perform on line computation scalability algorithms must have processing times that scale well with ever growing volumes automation the analysis process must be automated so that information extraction can run on its own ease of use everyday citizens should be able to extract and assess the necessary information and robustness systems must be able to cope with data of poor quality missing or erroneous data the nato advanced study institute asi on mining massive data sets for security held in italy september 2007 brought together around ninety participants to discuss these issues this publication includes the most important contributions but can of course not entirely reflect the lively interactions which allowed the participants to exchange their views and share their experience the bridge between academic methods and industrial constraints is systematically discussed throughout this volume will thus serve as a reference book for anyone interested in understanding the techniques for handling very large data sets and how to apply them in conjunction for solving security issues

Real-Time Data Mining 2009-08 computational intelligence data mining data stream mining machine vision pattern recognition

Data Mining for Geoinformatics 2013-08-16 internet usage has become a normal and essential aspect of everyday life due to the immense amount of information available on the web it has become obligatory to sift through and categorize the overload of data while removing redundant material collaborative filtering using data mining and analysis evaluates the latest patterns and trending topics in the utilization of data mining tools and filtering practices featuring emergent research and optimization techniques in the areas of opinion mining text mining and sentiment analysis as well as their various applications this book is an essential reference source for researchers and engineers interested in collaborative filtering

Pocket Data Mining 2013-10-19 data stream mining is becoming very important in many application areas such as the stock market network traffic web logs and atm transactions a data stream consists of an ordered sequence of instances and because there are usually a large number of instances along with limited computing and storage capabilities algorithms that read the data only once are preferred there has been some research that focuses on finding when a concept has changed given some knowledge about the previous instances in the data stream but little on determining the characteristics of that change in this thesis we concentrate on finding the characteristics of the changes that occur using frequent itemset mining techniques we propose two approaches both combining heuristic and statistical approaches to analyse changes that have occurred within a stream at itemset level hoeffding bound and bernstein bound inequality are two statistical methods used in our research in order to handle the infinite length and limited labelled data streams our approaches incorporate a sliding window algorithm and a couple of effective data structure such as cp tree can tree and spo tree to identify three types of change extension reduction and support fluctuation to experiment and evaluate our algorithms we customized the ibm quest synthetic data generator to create types of dataset containing those three types of change our algorithms have been fully tested on both synthetic and real world datasets

Heterogeneous Computing for Data Stream Mining 2016 data mining and machine learning applications the book elaborates in detail on the current needs of data mining and machine learning and promotes mutual understanding among research in different disciplines thus facilitating research development and collaboration data the latest currency of today s world is the new gold in this new form of gold the most beautiful jewels are data analytics and machine learning data mining and machine learning are considered interdisciplinary fields data
mining is a subset of data analytics and machine learning involves the use of algorithms that automatically improve through experience based on data massive datasets can be classified and clustered to obtain accurate results the most common technologies used include classification and clustering methods accuracy and error rates are calculated for regression and classification and clustering to find actual results through algorithms like support vector machines and neural networks with forward and backward propagation applications include fraud detection image processing medical diagnosis weather prediction e commerce and so forth the book features a review of the state of the art in data mining and machine learning a review and description of the learning methods in human computer interaction implementation strategies and future research directions used to meet the design and application requirements of several modern and real time applications for a long time the scope and implementation of a majority of data mining and machine learning strategies a discussion of real time problems audience industry and academic researchers scientists and engineers in information technology data science and machine and deep learning as well as artificial intelligence more broadly
Scientific Data Mining and Knowledge Discovery 2009-09-19 computational intelligence data mining data stream mining machine vision pattern recognition big data data science
Mining Massive Data Sets for Security 2008
Data Stream Mining & Processing (DSMP), IEEE International Conference on 2016-08-23
2016 IEEE First International Conference on Data Stream Mining and Processing (DSMP) 2016-07-13
Collaborative Filtering Using Data Mining and Analysis 2014
Change Detection in Pattern Stream Mining 2013
Proceedings of the 2016 IEEE First International Conference on Data Stream Mining & Processing (DSMP) 2022-01-26
Data Stream Mining with Multiple Sliding Windows for Continuous Prediction 2022
Data Mining and Machine Learning Applications 2002
Evolutionary-EAC Instance-learning-based Predictive Algorithm for Fast Data Stream Mining 2009
Time-critical Data Stream Mining for Mobile Environment Using MobiMINE 2011
Situation and Resource-aware Adaptation for Ubiquitous Data Stream Mining 2024
Dynamic Data Mining 2018-08-21
Data Stream Mining and Analysis
2018 IEEE Second International Conference on Data Stream Mining and Processing (DSMP)