Topological Data Analysis And Machine Learning Theory

Topological data analysis (TDA) has emerged recently as a viable tool for analyzing complex data, and the area has grown substantially both in its methodologies and applicability. Providing a computational and algorithmic foundation for techniques in TDA, this comprehensive, self-contained text introduces students and researchers in mathematics and computer science to the current state of the field. The book features a description of mathematical objects and constructs behind recent advances, the algorithms involved, computational considerations, as well as examples of topological structures or ideas that can be used in applications. It provides a thorough treatment of persistent homology together with various extensions - like zigzag persistence and multiparameter persistence - and their applications to different types of data, like point clouds, triangulations, or graph data. Other important topics covered include discrete Morse theory, the Mapper structure, optimal generating cycles, as well as recent advances in embedding TDA within machine learning frameworks.
Persistence theory emerged in the early 2000s as a new theory in the area of

applied and computational topology. This book provides a broad and modern view of the subject, including its algebraic, topological, and algorithmic aspects. It also elaborates on applications in data analysis. The level of detail of the exposition has been set so as to keep a survey style, while providing sufficient insights into the proofs so the reader can understand the mechanisms at work. The book is organized into three parts. The first part is dedicated to the foundations of persistence and emphasizes its connection to quiver representation theory. The second part focuses on its connection to applications through a few selected topics. The third part provides perspectives for both the theory and its applications. The book can be used as a text for a course on applied topology or data analysis.

One of the grand challenges in our digital world are the large, complex and often weakly structured data sets, and massive amounts of unstructured information. This “big data” challenge is most evident in biomedical informatics: the trend towards precision medicine has resulted in an explosion in the amount of generated biomedical data sets. Despite the fact that human experts are very good at pattern recognition in dimensions of = 3; most of the data is high-dimensional, which makes manual analysis often impossible and neither the medical doctor nor the biomedical researcher can memorize all these facts. A

synergistic combination of methodologies and approaches of two fields offer ideal conditions towards unraveling these problems: Human–Computer Interaction (HCI) and Knowledge Discovery/Data Mining (KDD), with the goal of supporting human capabilities with machine learning./ppThis state-of-the-art survey is an output of the HCI-KDD expert network and features 19 carefully selected and reviewed papers related to seven hot and promising research areas: Area 1: Data Integration, Data Pre-processing and Data Mapping; Area 2: Data Mining Algorithms; Area 3: Graph-based Data Mining; Area 4: Entropy-Based Data Mining; Area 5: Topological Data Mining; Area 6 Data Visualization and Area 7: Privacy, Data Protection, Safety and Security.

This collection of peer-reviewed workshop papers provides comprehensive coverage of cutting-edge research into topological approaches to data analysis and visualization. It encompasses the full range of new algorithms and insights, including fast homology computation, comparative analysis of simplification techniques, and key applications in materials and medical science. The book also addresses core research challenges such as the representation of large and complex datasets, and integrating numerical methods with robust combinatorial algorithms. In keeping with the focus of the TopoInVis 2017 Workshop, the contributions reflect the latest advances in finding experimental solutions to open

problems in the sector. They provide an essential snapshot of state-of-the-art research, helping researchers to keep abreast of the latest developments and providing a basis for future work. Gathering papers by some of the world's leading experts on topological techniques, the book represents a valuable contribution to a field of growing importance, with applications in disciplines ranging from engineering to medicine.

This book describes new theories and applications of artificial neural networks, with a special focus on neural computation, cognitive science and machine learning. It discusses cutting-edge research at the intersection between different fields, from topics such as cognition and behavior, motivation and emotions, to neurocomputing, deep learning, classification and clustering. Further topics include signal processing methods, robotics and neurobionics, and computer vision alike. The book includes selected papers from the XIX International Conference on Neuroinformatics, held on October 2-6, 2017, in Moscow, Russia.

This book describes current problems in data science and Big Data. Key topics are data classification, Graph Cut, the Laplacian Matrix, Google Page Rank, efficient algorithms, hardness of problems, different types of big data, geometric data structures, topological data processing, and various learning methods. For unsolved problems such as incomplete data relation and reconstruction, the book

includes possible solutions and both statistical and computational methods for data analysis. Initial chapters focus on exploring the properties of incomplete data sets and partial-connectedness among data points or data sets. Discussions also cover the completion problem of Netflix matrix; machine learning method on massive data sets; image segmentation and video search. This book introduces software tools for data science and Big Data such MapReduce, Hadoop, and Spark. This book contains three parts. The first part explores the fundamental tools of data science. It includes basic graph theoretical methods, statistical and AI methods for massive data sets. In second part, chapters focus on the procedural treatment of data science problems including machine learning methods, mathematical image and video processing, topological data analysis, and statistical methods. The final section provides case studies on special topics in variational learning, manifold learning, business and financial data rec overy, geometric search, and computing models. Mathematical Problems in Data Science is a valuable resource for researchers and professionals working in data science, information systems and networks. Advanced-level students studying computer science, electrical engineering and mathematics will also find the content helpful.

Topological data analysis (TDA) is a rapidly developing collection of methods for

studying the shape of data. Persistent homology is a prominent branch of TDA which analyzes the dynamics of topological features of a data set. We introduce statistical inference and learning methods for persistent homology of three types of data: point clouds, fingerprints, and rock images. First, we illustrate a topological inference plot for point cloud data, called the persistence terrace. The suggested plot allows robust and scale-free inference on the size and point density of topological features. Second, we suggest a new interface between persistent homology and machine learning algorithms and apply it to the problem of sorting fingerprints into pre-determined groups. We achieve near state-of-the-art classification accuracy rates by applying TDA to minutiae points and ink-roll images. Last, we present a statistical model for analysis of porous materials using persistent homology. Our model enables us to predict the geophysical properties of rocks based on their geometry and connectivity.

The volume of data is not the only problem in modern data analysis, data complexity is often more challenging. In many areas such as computational biology, topological data analysis, and machine learning, the data resides in high dimensional spaces which may not even be Euclidean. Therefore, processing such massive and complex data and extracting some useful information is a big challenge. Our methods will apply to any data sets given as a set of objects and

a metric that measures the distance between them. In this dissertation, we first consider the problem of preprocessing and organizing such complex data into a hierarchical data structure that allows efficient nearest neighbor and range queries. There have been many data structures for general metric spaces, but almost all of them have construction time that can be quadratic in terms of the number of points. There are only two data structures with O(n log n) construction time, but both have very complex algorithms and analyses. Also, they cannot be implemented efficiently. Here, we present a simple, randomized incremental algorithm that builds a metric data structure in O(n log n) time in expectation. Thus, we achieve the best of both worlds, simple implementation with asymptotically optimal performance. Furthermore, we consider the close relationship between our metric data structure and point orderings used in applications such as k-center clustering. We give linear time algorithms to go back and forth between these orderings and our metric data structure. In the last part, we use metric data structures to extract topological features of a data set, such as the number of connected components, holes, and voids. We give an efficient algorithm for constructing a (1 + epsilon)-approximation to the so-called Nerve filtration of a metric space, a fundamental tool in topological data analysis. This dissertation presents novel approaches and applications of machine

learning architectures. In particular, these approaches are based on tools from topological data analysis and are used in conjunction with conventional machine learning methods. Topological data analysis, which is based on algebraic topology, can identify significant global mathematical structures which are out of reach of many other approaches. When we use topology we benefit from generality, and when we use conventional methods we benefit from specificity. This dissertation contains a broad overview of data science and topological data analysis, then transitions to three distinct machine learning applications of these methods. The first application uses linear methods to discover the inherent dimensionality of the manifold given by congressional roll call votes. The second uses persistent homology to identify extremely noisy images in both supervised and unsupervised tasks. The last application uses mapper objects to produce robust classification algorithms. Two additional projects are presented later in the appendix, and are related to the three main applications. The first of these constructs a method to choose optimal optimizers, and the second places mathematical constraints on the structure of renormalization group flows. Advances in Nonlinear Geosciences is a set of contributions from the participants of “30 Years of Nonlinear Dynamics” held July 3-8, 2016 in Rhodes, Greece as part of the Aegean Conferences, as well as from several other experts in the field

who could not attend the meeting. The volume brings together up-to-date research from the atmospheric sciences, hydrology, geology, and other areas of geosciences and presents the new advances made in the last 10 years. Topics include chaos synchronization, topological data analysis, new insights on fractals, multifractals and stochasticity, climate dynamics, extreme events, complexity, and causality, among other topics.

This will be a comprehensive, multi-contributed reference work that will detail the latest research and developments in biomedical signal processing related to big data medical analysis. It will describe signal processing, machine learning, and parallel computing strategies to revolutionize the world of medical analytics and diagnosis as presented by world class researchers and experts in this important field. The chapters will desribe tools that can be used by biomedical and clinical practitioners as well as industry professionals. It will give signal processing researchers a glimpse into the issues faced with Big Medical Data.

Image Processing, Analysis and Machine Vision represent an exciting part of modern cognitive and computer science. Following an explosion of inter est during the Seventies, the Eighties were characterized by the maturing of the field and the significant growth of active applications; Remote Sensing, Technical Diagnostics, Autonomous Vehicle Guidance and Medical Imaging are the most

rapidly developing areas. This progress can be seen in an in creasing number of software and hardware products on the market as well as in a number of digital image processing and machine vision courses offered at universities world-wide. There are many texts available in the areas we cover - most (indeed, all of which we know) are referenced somewhere in this book. The subject suffers, however, from a shortage of texts at the 'elementary' level - that appropriate for undergraduates beginning or completing their studies of the topic, or for Master's students - and the very rapid developments that have taken and are still taking place, which quickly age some of the very good text books produced over the last decade or so. This book reflects the authors' experience in teaching one and two semester undergraduate and graduate courses in Digital Image Processing, Digital Image Analysis, Machine Vision, Pattern Recognition and Intelligent Robotics at their respective institutions.

This book features selected high-quality research papers presented at the International Conference on Machine Intelligence and Signal Processing (MISP 2019), held at the Indian Institute of Technology, Allahabad, India, on September 7–10, 2019. The book covers the latest advances in the fields of machine learning, big data analytics, signal processing, computational learning theory, and their real-time applications. The topics covered include support vector machines

(SVM) and variants like least-squares SVM (LS-SVM) and twin SVM (TWSVM), extreme learning machine (ELM), artificial neural network (ANN), and other areas in machine learning. Further, it discusses the real-time challenges involved in processing big data and adapting the algorithms dynamically to improve the computational efficiency. Lastly, it describes recent developments in processing signals, for instance, signals generated from IoT devices, smart systems, speech, and videos and addresses biomedical signal processing: electrocardiogram (ECG) and electroencephalogram (EEG).

This timely text introduces topological data analysis from scratch, with detailed case studies.

This book constitutes the refereed proceedings of the 5th International Workshop on Machine Learning in Medical Imaging, MLMI 2014, held in conjunction with the International Conference on Medical Image Computing and Computer Assisted Intervention, MICCAI 2014, in Cambridge, MA, USA, in September 2014. The 40 contributions included in this volume were carefully reviewed and selected from 70 submissions. They focus on major trends and challenges in the area of machine learning in medical imaging and aim to identify new cutting-edge techniques and their use in medical imaging.

An introduction to geometric and topological methods to analyze large scale biological data; includes statistics and genomic applications.

This coherent mathematical and statistical approach aimed at graduate students incorporates

regression and topology as well as graph theory.

Over the past years, businesses have had to tackle the issues caused by numerous forces from political, technological and societal environment. The changes in the global market and increasing uncertainty require us to focus on disruptive innovations and to investigate this phenomenon from different perspectives. The benefits of innovations are related to lower costs, improved efficiency, reduced risk, and better response to the customers' needs due to new products, services or processes. On the other hand, new business models expose various risks, such as cyber risks, operational risks, regulatory risks, and others. Therefore, we believe that the entrepreneurial behavior and global mindset of decision-makers significantly contribute to the development of innovations, which benefit by closing the prevailing gap between developed and developing countries. Thus, this Special Issue contributes to closing the research gap in the literature by providing a platform for a scientific debate on innovation, internationalization and entrepreneurship, which would facilitate improving the resilience of businesses to future disruptions. Order Your Print Copy

A rigorous introduction to geometric and topological inference, for anyone interested in a geometric approach to data science.

In Geography and GIS, surfaces can be analysed and visualised through various data structures, and topological data structures describe surfaces in the form of a relationship between certain surface-specific features. Drawn from many disciplines with a strong applied aspect, this is a research-led, interdisciplinary approach to the creation, analysis and visualisation of surfaces, focussing on topological data structures. Topological Data Structures for Surfaces: an introduction for Geographical Information Science describes the concepts and

applications of these data structures. The book focuses on how these data structures can be used to analyse and visualise surface datasets from a range of disciplines such as human geography, computer graphics, metrology, and physical geography. Divided into two Parts, Part I defines the topological surface data structures and explains the various automated methods used for their generation. Part II demonstrates a number of applications of surface networks in diverse fields, ranging from sub-atomic particle collision visualisation to the study of population density patterns. To ensure that the material is accessible, each Part is prefaced by an overview of the techniques and application. Provides GI scientists and geographers with an accessible overview of current surface topology research. Algorithms are presented and explained with practical examples of their usage. Features an accompanying website developed by the Editor - http://geog.le.ac.uk/sanjayrana/surface-networks/ This book is invaluable for researchers and postgraduate students working in departments of GI Science, Geography and Computer Science. It also constitutes key reference material for Masters students working on surface analysis projects as part of a GI Science or Computer Science programme.

This book presents the proceedings of the 24th European Conference on Artificial Intelligence (ECAI 2020), held in Santiago de Compostela, Spain, from 29 August to 8 September 2020. The conference was postponed from June, and much of it conducted online due to the COVID-19 restrictions. The conference is one of the principal occasions for researchers and practitioners of AI to meet and discuss the latest trends and challenges in all fields of AI and to demonstrate innovative applications and uses of advanced AI technology. The book also includes the proceedings of the 10th Conference on Prestigious Applications of Artificial

Intelligence (PAIS 2020) held at the same time. A record number of more than 1,700 submissions was received for ECAI 2020, of which 1,443 were reviewed. Of these, 361 full-papers and 36 highlight papers were accepted (an acceptance rate of 25% for full-papers and 45% for highlight papers). The book is divided into three sections: ECAI full papers; ECAI highlight papers; and PAIS papers. The topics of these papers cover all aspects of AI, including Agent-based and Multi-agent Systems; Computational Intelligence; Constraints and Satisfiability; Games and Virtual Environments; Heuristic Search; Human Aspects in AI; Information Retrieval and Filtering; Knowledge Representation and Reasoning; Machine Learning; Multidisciplinary Topics and Applications; Natural Language Processing; Planning and Scheduling; Robotics; Safe, Explainable, and Trustworthy AI; Semantic Technologies; Uncertainty in AI; and Vision. The book will be of interest to all those whose work involves the use of AI technology.

Combining theoretical and practical aspects of topology, this book provides a comprehensive and self-contained introduction to topological methods for the analysis and visualization of scientific data. Theoretical concepts are presented in a painstaking but intuitive manner, with numerous high-quality color illustrations. Key algorithms for the computation and simplification of topological data representations are described in detail, and their application is carefully demonstrated in a chapter dedicated to concrete use cases. With its fine balance between theory and practice, "Topological Data Analysis for Scientific Visualization" constitutes an appealing introduction to the increasingly important topic of topological data analysis for lecturers, students and researchers.

The three volume proceedings LNAI 11051 – 11053 constitutes the refereed proceedings of

the European Conference on Machine Learning and Knowledge Discovery in Databases, ECML PKDD 2018, held in Dublin, Ireland, in September 2018. The total of 131 regular papers presented in part I and part II was carefully reviewed and selected from 535 submissions; there are 52 papers in the applied data science, nectar and demo track. The contributions were organized in topical sections named as follows: Part I: adversarial learning; anomaly and outlier detection; applications; classification; clustering and unsupervised learning; deep learningensemble methods; and evaluation. Part II: graphs; kernel methods; learning paradigms; matrix and tensor analysis; online and active learning; pattern and sequence mining; probabilistic models and statistical methods; recommender systems; and transfer learning. Part III: ADS data science applications; ADS e-commerce; ADS engineering and design; ADS financial and security; ADS health; ADS sensing and positioning; nectar track; and demo track.

The BIRS Workshop “Advances in Interactive Knowledge Discovery and Data Mining in Complex and Big Data Sets” (15w2181), held in July 2015 in Banff, Canada, was dedicated to stimulating a cross-domain integrative machine-learning approach and appraisal of “hot topics” toward tackling the grand challenge of reaching a level of useful and useable computational intelligence with a focus on real-world problems, such as in the health domain. This encompasses learning from prior data, extracting and discovering knowledge, generalizing the results, fighting the curse of dimensionality, and ultimately disentangling the underlying explanatory factors in complex data, i.e., to make sense of data within the context of the application domain. The workshop aimed to contribute advancements in promising novel areas such as at the intersection of machine learning and topological data analysis. History

has shown that most often the overlapping areas at intersections of seemingly disparate fields are key for the stimulation of new insights and further advances. This is particularly true for the extremely broad field of machine learning.

This book constitutes the refereed joint proceedings of the 4th International Workshop on Interpretability of Machine Intelligence in Medical Image Computing, iMIMIC 2020, and the First International Workshop on Topological Data Analysis and Its Applications for Medical Data, TDA4MedicalData 2021, held on September 27, 2021, in conjunction with the 24th International Conference on Medical Imaging and Computer-Assisted Intervention, MICCAI 2021. The 7 full papers presented at iMIMIC 2021 and 5 full papers held at TDA4MedicalData 2021 were carefully reviewed and selected from 12 submissions each. The iMIMIC papers focus on introducing the challenges and opportunities related to the topic of interpretability of machine learning systems in the context of medical imaging and computer assisted intervention. TDA4MedicalData is focusing on using TDA techniques to enhance the performance, generalizability, efficiency, and explainability of the current methods applied to medical data.

This book constitutes the refereed proceedings of the IFIP TC 5, TC 12, WG 8.4, 8.9, 12.9 International Cross-Domain Conference for Machine Learning and Knowledge Extraction, CD-MAKE 2019, held in Canterbury, UK, in August 2019. The 25 revised full papers presented were carefully reviewed and selected from 45 submissions. The cross-domain integration and appraisal of different fields provides an atmosphere to foster different perspectives and opinions; it will offer a platform for novel ideas and a fresh look on the methodologies to put these ideas into business for the benefit of humanity.

Many real-world data sets can be viewed as a noisy sampling of an unknown high-dimensional topological space. The emergence and development of topological data analysis (TDA) over the last fifteen years or so provides a suite of tools to understand and exploit the topological structure of the underlying space from a multi-scale perspective that characterizes the shape of the data. This dissertation, thus, aims to leverage the shape information of data offered by the TDA tools to extract key features in machine learning and network science problems. We investigate a few TDA topics that are understudied following this line of research. We first extend the application of TDA to the manufacturing systems domain. We apply a widely used TDA method, known as the Mapper algorithm, on two benchmark data sets for chemical process yield prediction and semiconductor wafer fault detection. The algorithm yields topological networks that capture the intrinsic clusters and connections among the clusters (i.e., subgroups) present in the data sets, which are difficult to detect using traditional methods. Key process variables (features) that best differentiate the subgroups of interest are subsequently identified through statistical tests. Next we present a new method, referred as Sparse-TDA method, that integrates QR pivoting-based sparse sampling algorithm into vector-based TDA method to transform topological features into image pixels and identify discriminative pixel samples (features) in the presence of noisy and redundant information. We demonstrate its advantage over a state- of-the-art kernel TDA method and L1-regularized feature selection methods in terms of classification accuracy and training time on three challenging data sets pertaining to 3D meshes of synthetic and real human postures and textured images. Finally, we propose a method that extends the persistence-based TDA that is typically used for characterizing shapes to general networks. We introduce the concept of the

community tree, a tree structure established based on clique communities from the clique percolation method, to summarize the topological structures in a network from a persistence perspective. Furthermore, we develop efficient algorithms to construct and update community trees by maintaining a series of clique graphs in the form of spanning forests, in which each spanning tree is built on an underlying Euler Tour tree. With the information revealed by community trees and the corresponding persistence diagrams, our proposed approach is able to detect clique communities and keep track of the major structural changes during their evolution given a stability threshold. The results demonstrate its effectiveness in extracting useful structural insights for time-varying social networks.

This book gathers the proceedings of the 2018 Abel Symposium, which was held in Geiranger, Norway, on June 4-8, 2018. The symposium offered an overview of the emerging field of "Topological Data Analysis". This volume presents papers on various research directions, notably including applications in neuroscience, materials science, cancer biology, and immune response. Providing an essential snapshot of the status quo, it represents a valuable asset for practitioners and those considering entering the field.

Machine Intelligence for Healthcare is a must read for physician leaders, health insurance executives, clinical researchers, public health officials, data scientists and software engineers seeking to understand this pivotal innovation in the information revolution in healthcare. MI for Healthcare provides a detailed introduction of Machine Intelligence, then takes the reader on a journey through the basics of machine learning, topological data analysis and applications of machine intelligence software for

healthcare and life sciences. Over 20 case studies cover topics related to clinical variation analysis, hospital clinical pathways, population health management, genetic analysis, precision medicine, healthcare revenue cycle, and payment integrity. The book includes a detailed introduction of the mathematics of topology and concepts of machine learning algorithms. This provides an understanding for the central role which machine intelligence software is now playing in the emergence of the "learning healthcare system" and success in the new world of value-based healthcare delivery. Social machines are a type of network connected by interactive digital devices made possible by the ubiquitous adoption of technologies such as the Internet, the smartphone, social media and the read/write World Wide Web, connecting people at scale to document situations, cooperate on tasks, exchange information, or even simply to play. Existing social processes may be scaled up, and new social processes enabled, to solve problems, augment reality, create new sources of value, and disrupt existing practice. This book considers what talents one would need to understand or build a social machine, describes the state of the art, and speculates on the future, from the perspective of the EPSRC project SOCIAM – The Theory and Practice of Social Machines. The aim is to develop a set of tools and techniques for investigating, constructing and facilitating social machines, to enable us to narrow down pragmatically what is becoming a wide space, by asking ‘when will it be valuable to use these methods on a sociotechnical system?’ The systems for which the use of these

methods adds value are social machines in which there is rich person-to-person communication, and where a large proportion of the machine's behaviour is constituted by human interaction.

Integration of Topological Data Analysis and Machine Learning for Small Molecule Property PredictionsInterpretability of Machine Intelligence in Medical Image Computing, and Topological Data Analysis and Its Applications for Medical Data4th International Workshop, iMIMIC 2021, and 1st International Workshop, TDA4MedicalData 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, ProceedingsSpringer Nature

This 8-volumes set constitutes the refereed of the 25th International Conference on Pattern Recognition Workshops, ICPR 2020, held virtually in Milan, Italy and rescheduled to January 10 - 11, 2021 due to Covid-19 pandemic. The 416 full papers presented in these 8 volumes were carefully reviewed and selected from about 700 submissions. The 46 workshops cover a wide range of areas including machine learning, pattern analysis, healthcare, human behavior, environment, surveillance, forensics and biometrics, robotics and egovision, cultural heritage and document analysis, retrieval, and women at ICPR2020.

The multi-volume set LNAI 12975 until 12979 constitutes the refereed proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases, ECML PKDD 2021, which was held during September 13-17, 2021. The

conference was originally planned to take place in Bilbao, Spain, but changed to an online event due to the COVID-19 pandemic. The 210 full papers presented in these proceedings were carefully reviewed and selected from a total of 869 submissions. The volumes are organized in topical sections as follows: Research Track: Part I: Online learning; reinforcement learning; time series, streams, and sequence models; transfer and multi-task learning; semi-supervised and few-shot learning; learning algorithms and applications. Part II: Generative models; algorithms and learning theory; graphs and networks; interpretation, explainability, transparency, safety. Part III: Generative models; search and optimization; supervised learning; text mining and natural language processing; image processing, computer vision and visual analytics. Applied Data Science Track: Part IV: Anomaly detection and malware; spatio-temporal data; e-commerce and finance; healthcare and medical applications (including Covid); mobility and transportation. Part V: Automating machine learning, optimization, and feature engineering; machine learning based simulations and knowledge discovery; recommender systems and behavior modeling; natural language processing; remote sensing, image and video processing; social media.