Computer Science

Chapman & Hall/CRC
Data Mining and Knowledge Discovery Series

Advances in Machine Learning and Data Mining for Astronomy

The First Comprehensive Book on Data Mining and Machine Learning Techniques and Tools for Astronomy Applications

Advances in Machine Learning and Data Mining for Astronomy documents numerous successful collaborations among computer scientists, statisticians, and astronomers who illustrate the application of state-of-the-art machine learning and data mining techniques in astronomy. Due to the massive amount and complexity of data in most scientific disciplines, the material discussed in this text transcends traditional boundaries between various areas in the sciences and computer science.

The book's introductory part provides context to issues in the astronomical sciences that are also important to health, social, and physical sciences, particularly probabilistic and statistical aspects of classification and cluster analysis. The next part describes a number of astrophysics case studies that leverage a range of machine learning and data mining technologies. In the last part, developers of algorithms and practitioners of machine learning and data mining show how these tools and techniques are used in astronomical applications.

With contributions from leading astronomers and computer scientists, this book is a practical guide to many of the most important developments in machine learning, data mining, and statistics. It explores how these advances can solve current and future problems in astronomy and looks at how they could lead to the creation of entirely new algorithms within the data mining community.

ISBN 978-1-138-19930-9

CRC Press
Taylor & Francis Group
an **informa** business
www.crcpress.com

Foreword, ix

Editors, xi

Perspective, xiii

Contributors, xxv

PART I **Foundational Issues**

CHAPTER 1 ■ Classification in Astronomy: Past and Present 3

ERIC FEIGELSON

CHAPTER 2 ■ Searching the Heavens: Astronomy, Computation, Statistics, Data Mining, and Philosophy 11

CLARK GLYMOUR

CHAPTER 3 ■ Probability and Statistics in Astronomical Machine Learning and Data Mining 27

JEFFREY D. SCARGLE

PART II **Astronomical Applications**

SECTION 1 **Source Identification**

CHAPTER 4 ■ Automated Science Processing for the Fermi Large Area Telescope 41

JAMES CHIANG

CHAPTER 5 ■ Cosmic Microwave Background Data Analysis 55

PANIEZ PAYKARI AND JEAN-LUC STARCK

CHAPTER 6 ■ Data Mining and Machine Learning in Time-Domain Discovery and Classification 89

JOSHUA S. BLOOM AND JOSEPH W. RICHARDS

CHAPTER 7 ■ Cross-Identification of Sources: Theory and Practice 113

TAMÁS BUDAVÁRI

CHAPTER 8 ■ The Sky Pixelization for Cosmic Microwave Background Mapping 133

O.V. VERKHODANOV AND A.G. DOROSHKEVICH

CHAPTER 9 ■ Future Sky Surveys: New Discovery Frontiers 161

J. ANTHONY TYSON AND KIRK D. BORNE

CHAPTER 10 ■ Poisson Noise Removal in Spherical Multichannel Images: Application to Fermi Data 183

JÉRÉMY SCHMITT, JEAN-LUC STARCK, JALAL FADILI, AND SETH DIGEL

SECTION 2 **Classification**

CHAPTER 11 ■ Galaxy Zoo: Morphological Classification and Citizen Science 213

LUCY FORTSON, KAREN MASTERS, ROBERT NICHOL, KIRK D. BORNE, EDWARD M. EDMONDSON, CHRIS LINTOTT, JORDAN RADDICK, KEVIN SCHAWINSKI, AND JOHN WALLIN

CHAPTER 12 ■ The Utilization of Classifications in High-Energy Astrophysics Experiments 237

BILL ATWOOD

CHAPTER 13 ■ Database-Driven Analyses of Astronomical Spectra 267

JAN CAMI

CHAPTER 14 ■ Weak Gravitational Lensing 287

SANDRINE PIRES, JEAN-LUC STARCK, ADRIENNE LEONARD, AND ALEXANDRE RÉFRÉGIER

CHAPTER 15 ■ Photometric Redshifts: 50 Years After 323

TAMÁS BUDAVÁRI

CHAPTER 16 ■ Galaxy Clusters 337

CHRISTOPHER J. MILLER

SECTION 3 **Signal Processing (Time-Series) Analysis**

CHAPTER 17 ■ Planet Detection: The Kepler Mission 355

JON M. JENKINS, JEFFREY C. SMITH, PETER TENENBAUM, JOSEPH D. TWICKEN, AND JEFFREY VAN CLEVE

CHAPTER 18 ■ Classification of Variable Objects in Massive Sky Monitoring Surveys 383

PRZEMEK WOŹNIAK, ŁUKASZ WYRZYKOWSKI, AND VASILY BELOKUROV

CHAPTER 19 ■ Gravitational Wave Astronomy 407

LEE SAMUEL FINN

SECTION 4 **The Largest Data Sets**

CHAPTER 20 ■ Virtual Observatory and Distributed Data Mining 447

KIRK D. BORNE

CHAPTER 21 ■ Multitree Algorithms for Large-Scale Astrostatistics 463

WILLIAM B. MARCH, ARKADAS OZAKIN, DONGRYEOL LEE, RYAN RIEGEL, AND ALEXANDER G. GRAY

PART III **Machine Learning Methods**

CHAPTER 22 ■ Time–Frequency Learning Machines for Nonstationarity Detection Using Surrogates 487

PIERRE BORGNAT, PATRICK FLANDRIN, CÉDRIC RICHARD, ANDRÉ FERRARI, HASSAN AMOUD, AND PAUL HONEINE

CHAPTER 23 ■ Classification 505

NIKUNJ OZA

CHAPTER 24 ■ On the Shoulders of Gauss, Bessel, and Poisson: Links, Chunks, Spheres, and Conditional Models 523

WILLIAM D. HEAVLIN

CHAPTER 25 ■ Data Clustering 543

KIRI L. WAGSTAFF

CHAPTER 26 ■ Ensemble Methods: A Review 563

MATTEO RE AND GIORGIO VALENTINI

CHAPTER 27 ■ Parallel and Distributed Data Mining for Astronomy Applications 595

KAMALIKA DAS AND KANISHKA BHADURI

CHAPTER 28 ■ Pattern Recognition in Time Series 617

JESSICA LIN, SHERI WILLIAMSON, KIRK D. BORNE, AND DAVID DEBARR

CHAPTER 29 ■ Randomized Algorithms for Matrices and Data 647

MICHAEL W. MAHONEY

INDEX, 673